DOCTOR OF PHILOSOPHY

Proteomics and dynamics of the human nucleolus

Leung, Anthony Kar Lun

Award date:
2003

Awarding institution:
University of Dundee

DOCTOR OF PHILOSOPHY

Proteomics and dynamics of the human nucleolus

Anthony Kar Lun Leung

2003

University of Dundee

PhD Thesis

presented in partial fulfillment of the degree of

Doctor of Philosophy to the University of Dundee

PROTEOMICS AND DYNAMICS OF THE HUMAN NUCLEOLUS

Anthony Kar Lun Leung

School of Life Sciences

University of Dundee

Scotland

Supervisor: Prof. Angus I. Lamond

January 2003

Table of contents

List of Figures

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

List of Tables

Acknowledgement

I thank my mentor and supervisor Prof. Angus Lamond for his inspirations on the project and invaluable advices on many aspects of my life, especially when things do not seem to get anywhere. A lot of thanks go to Carol, Archa, Yun and Jen (University of Odense, Denmark) in developing and sharing the ideas for the ongoing nucleolar proteome project. Thanks for Carol, Yun and Alfred to provide nice office environment and Archa, Ursula and Steve in the bench work area. Thanks for Steve and Paul for many scientific advices and clear explanations. Thanks for Judith, Laura and Kevin for making the laboratory much more relaxed and could not imagine what it would be like without them. Thanks for Yun, Janet and Melpi for sharing the long hard night with music and food. Thanks for Diego for working on the follow-up of the bioinformatics analysis. Thanks for Angus, Archa, Alfred, Yun, Alice and Tunhan for comments on this manuscript and their meticulous eyes in spotting English errors written by a foreigner.

The project would not be possible without the financial support from the Croucher Foundation (Hong Kong, China) and Overseas Research Scheme Award (UK). Thanks for the European Advanced Light Microscopy Facilities short-term fellowship which made the visitation trip to use the state-of-art microscopy in the EMBL, Heidelberg possible. I thank Dr. Jan Ellenberg for hosting me there and giving advices on confocal microscopy and photobleaching analyses, and Nathalie, Daniel, Gwen, Peter, Joël and Katharina in his laboratory for sharing meals and ideas, especially in the little hours.

I thank Prof. Thoru Pederson (University of Massachusetts Medical School, USA) for sharing the story behind the discovery of the nucleolus, Dr. Bruno Frenguelli for the use of Zeiss LSM 510 confocal microscope in Ninewells Hospital, Profs. and Drs Geoff

Barton, Chris Proud, Joost Zomerdijk and Jason Swedlow for helpful discussion and Dr David Martin, Nick Helps and Sathya Shreeman for technical support.

Finally, I would like to take this opportunity to thank my mother, father and sister and also other family members for believing in me to make an unconventional choice to study PhD in Dundee and thank for their support which never seems far away.

"Lift your eyes and look to the heavens

(or after looking under the microscope):

Who created all these?

He who brings out the starry host one by one,

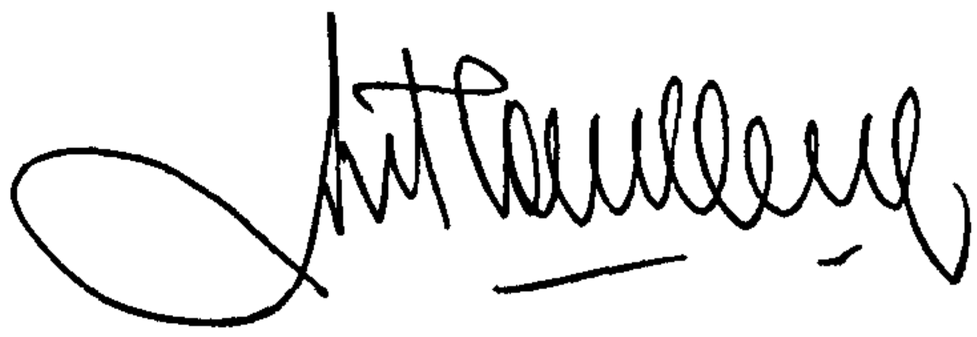
and call them each by name.

Because of his great power and mighty strength,

not one of them is missing."

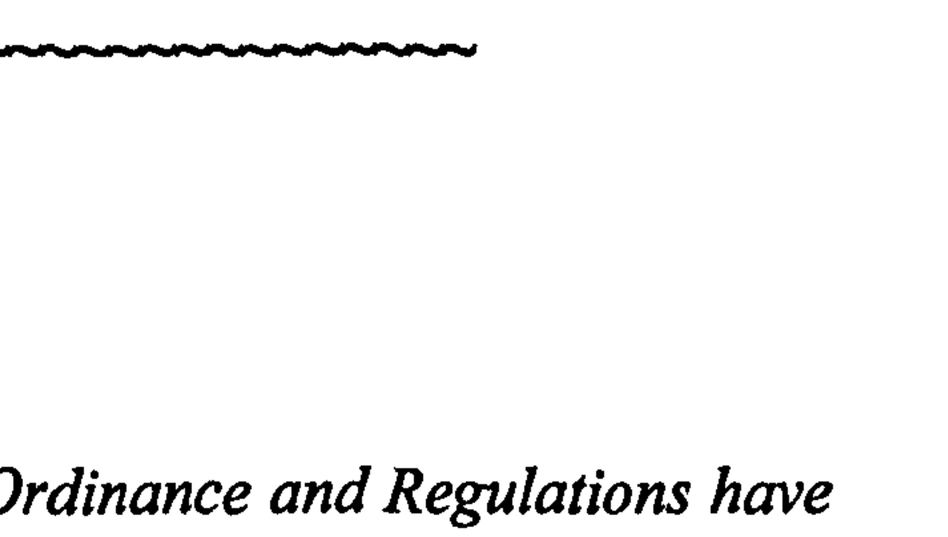
Declaration

I, hereby, declare that (a) I am the author of this thesis; (b) that unless otherwise stated, all references cited have been consulted; (c) that the work of this thesis is an account of my work during the PhD; and (d) that it has not been previously accepted for a higher degree.



Anthony Leung

I, hereby, declare that the conditions of the relevant Ordinance and Regulations have been fulfilled by this candidate.



Prof. Angus Lamond

Abstract

The nucleolus is the most prominent structure within the eukaryotic cell nucleus and it was established to be the site where the majority of ribosomal RNAs (5.8S, 18S and 28S) are transcribed, processed and assembled with ribosomal proteins to form ribosomal subunits. The sole role of ribosome biogenesis, however, cannot explain the specific nucleolar localisations of tumour suppressors, cell cycle-regulatory factors and viral proteins. Therefore, together with my colleagues in the laboratory of Prof. Angus Lamond, we carried out a proteomic approach with an aim to identify the core components of the human nucleolus isolated from HeLa cell nuclei. My role in this project includes verification of the newly identified components, database construction archiving the primary data and providing links to other related information in the public domain, and subsequent bioinformatics and microscopic analyses. So far, 400 proteins were identified in which ~30% represents novel or uncharacterised proteins, partly reflecting the current poor status in the human genome annotation, but also reflecting the unknown complexity of the nucleolus. To facilitate the understanding of the functions of these novel proteins, I used deposited data of their gene activities and homologues across the species to identify *in silico* those novel proteins that are likely to be involved in ribosomal biogenesis.

Like the nucleus, the nucleolus itself is subcompartmentalised into different domains, namely, the fibrillar centre, the dense fibrillar components and the granular components and these structures are disassembled and reassembled during mitosis in human cells. In order to understand the intricate mechanism behind these mitotic dynamics, I have generated a panel of 24 HeLa cell lines stably expressing one or more nucleolar marker to study the inter-relationships between these subnucleolar domains as well as their relationships with the chromosomes. The results suggest that

(1) a core subunit of the RNA polymerase I dissociates from the chromosomes between prophase and metaphase and (2) the breakdown and reassembly are dependent on the dissociation and the recruitment of RNA polymerase I to the chromosomes respectively.

As part of the follow-up to the nucleolar proteome identified, the study of one uncharacterised factor NHPX led to the discovery of a novel nucleolar targeting pathway that is observed in both primary and transformed cell lines. Although NHPX co-localises with the dense fibrillar component marker fibrillarin, NHPX transiently transits through the splicing speckles prior to the nucleolar accumulation whilst fibrillarin accumulates within the nucleolus immediately after the nuclear entry. The NHPX progression is dependent on pre-mRNA transcription and may link multiple RNA metabolic pathways that occur in distinct subnuclear domains.

Abbreviations

ActD	Actinomycin D
AMT tag	Accurate Mass Time tag
APL	Acute Promyelocytic Leukaemia
ATP	Adenosine TriPhosphate
BLAST	Basic Local Alignment Search Tool
bp	base pair
BrdUTP	BromodeoxyUridine TriPhosphate
BrUTP	BromoUridine TriPhosphate
BSA	Bovine Serum Albumin
CB	Cajal Body
CBP	CREB-Binding Protein
CCD	Charge-Coupled Device
CDD	Conserved Domain Database
cDNA	complementary DNA
CID spectra	Collision Induced Dissociation spectra
CSK buffer	CytoSKeletal buffer
Cy3dUTP	Cy3-deoxyUridine TriPhosphate
Da	Dalton
DAB	DiAminoBenzidine
DFC	Dense Fibrillar Component
DMEM	Dulbecco's Modified Eagle Medium
DMSO	DiMethyl SulphOxide
DNA	deoxyriboNucleic Acid
dNTP	DeoxyNucleotide TriPhosphate
DRB	5,6-dichloro-D-ribofuranosylbenzimidazole
DREAMS	Dynamic Range Enhancement Applied to Mass Spectrometry
DTT	DiThioThreitol
ECFP	Enhanced Cyan Fluorescent Protein
EDTA	EthyleneDiamine Tetra-Acetic acid
EGFP	Enhanced Green Fluorescent Protein
EGTA	1,2-Di (2-aminoethoxy) ethane-N, N, N'-tetra-acetic acid
EM	Electron Microscope
ER	Endoplasmic Reticulum
ESI	ElectroSpray Ionisation
EST	Expressed Sequence Tag
EtBr	Ethidium Bromide
EYFP	Enhanced Yellow Fluorescent Protein
f/c	final concentration
FC	Fibrillar Centre
FCS	Fetal Calf Serum
FCS	Fluorescence Correlation Spectroscopy
FISH	Fluorescence *In Situ* Hybridisation
FITC	Fluorescein IsoThioCyanate
FLIP	Fluorescence Loss In Photobleaching
FP	Fluorescent Protein
FRAP	Fluorescence Recovery After Photobleaching
FTIC	Fourier Transform Ion Cyclotron

G418	Geneticin
GC	Granular Component
GDP	Guanosine DiPhosphate
GTP	Guanosine TriPhosphate
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethane-sulphonic acid
HRP	HorseRadish Peroxidase
ICAT	Isotope Coded Affinity Tags
iFRAP	Inverse FRAP
IGC	Interchromatin Granule Clusters
IPI	International Protein Index
IT	Ion Trap
kbp	kilobasepair
kDa	kiloDalton
kV	kilovolt
LB	Luria-Bertani Medium
LC	Liquid Chromatography
LDS	Lithium Dodecyl Sulphate
LSM	Laser Scanning confocal Microscope
MALDI	Matrix-Assisted Laser Desorption Ionisation
MOPS	MOrpholinoPropane Sulphonic acid
MPM	MultiPhoton Microscope
MQ H_2O	Milli-Q H_2O
mRNA	messenger RNA
MS	Mass Spectrometry
Mw	Molecular weight
NLS	Nuclear Localisation Signal
NOR	Nucleolar Organising Region
NP40	Nonidet-P40
NPC	Nuclear Pore Complex
PAGE	PolyAcrylamide Gel Electrophoresis
PBS	Phosphate Buffer Saline
PBS-T	Phosphate Buffer Saline with Tween-20
PCR	Polymerase Chain Reaction
PEG	PolyEthylene Glycol
PF	Perichromatin Fibril
PFA	ParaFormAldehyde
pI	Isoeletric point
PIPES	Piperazine-N,N'-bis[2-ethane sulphonic acid]
PNB	PreNucleolar Body
PNC	PeriNucleolar Compartment
POD	PML Oncogenic Domain
PP1	Protein Phosphatase 1
ppm	part per million
PS	Penicillin-Streptomycin
PSF	Point Spread Function
PSP1	ParaSpeckle Protein 1
PTB	Polypyrimidine Tract Binding protein
QTOF	Quadrupole Time-Of-Flight
RB	RetinoBlastoma protein

RNA	RiboNucleic Acid
RNP	RiboNucleoProtein particle
rpm	revolutions per minute
rRNA	ribosomal RNA
SAGE	Serial Analysis of Gene Expression
scaRNA	small cajal body specific RNA
SDS	Sodium Dodecyl Sulphate
SIR family	Silent Information Regulatory family
snoRNA	small nucleolar RNA
snoRNP	small nucleolar RiboNucleoProtein particle
snRNA	small nuclear RNA
snRNP	small nuclear RiboNucleoProtein particle
SRP	Signal Recognition Particle
SUMO	*Small Ubiquitin-like MOdifier*
TAE	Tris Acetate EDTA buffer
TB	Terrific Broth
TBE	Tris Borate EDTA buffer
TEM	Transmission Electron Microscope
TOF	Time-Of-Flight
5' TOP	5' Terminal OligoPyrimidine
tRNA	transfer RNA
UV	UltraViolet
v/v	volume/volume
w/v	weight/volume
WFM	Wide Field Microscope

Chapter I

Introduction

"In jeder Zelle ohne Ausnahme befindet sich ein etwas dunklerer und compakter Nucleus von runder oder länglich runder Form. Er nimmt grösstenteils die Mitte einer jeden Zelle ein, besteht aus einem feinkörnigen Wesen, enthält aber in seinem Innern ein genau rundes Körperchen, welches auf diese Weise in ihm selbst wiederum eine Art von zweitem Nucleus bildet."

~ G. Valentin, 1836 ~

I. Introduction

I.1 Overview of Nuclear Architecture

I.1.1 General

Eukaryotic cells, by definition, keep their DNA in a separate internal compartment known as the nucleus. The observation of a nucleus was first documented on 3rd March, 1682 by Antoni van Leeuwenhoek (Van Leeuwenhoek, 1948). However, it was not until 1831 that the nucleus was accepted as a widespread constituent of different somatic cells (Brown, 1833). The nucleus is separated from the cytoplasm by the nuclear envelope, which consists of a double-layered membrane (Burke and Ellenberg, 2002; Holaska et al., 2002). To facilitate the transport of macromolecules, such as RNA and proteins, across the nuclear membrane, aqueous channels, known as nuclear pore complexes (NPC), allow passive diffusion of small molecules (up to 40-60 kDa) and active transport of larger macromolecules (Gorlich and Kutay, 1999; Rabut and Ellenberg, 2001; Weis, 2002). These large macromolecules require appropriate transport signals and receptors, known as importins and exportins, for nucleocytoplasmic transport (Nigg, 1997; Weis, 1998).

The nucleus of a typical human diploid cell is a spheroid with a volume of approximately 500μm^3 (Carmo-Fonseca et al., 1996). Within this space, there is approximately 6 x 10^9 base pair (bp) of DNA segregated into 23 sets of chromosome pairs. The DNA together with RNA and protein gives rise to a nuclear viscosity that is ~ six- to few hundred-fold higher than the cytoplasm (Lang et al., 1986; Alexander and Rieder, 1991). However, this double-membraned compartment is not a homogenous soup of macromolecules (Moneron and Bernhard, 1969). Rather, by the early 1800s, it was already obvious from the light microscope that the nuclear interior was not uniform, most notably with the discovery of 'a secondary nucleus within the first one', i.e. the nucleolus (Section I.2).

The landscape within the nucleus was further charted with the arrivals of the electron microscope and the fluorescence microscope in the 1960s and by employing different dyes and fixation methods to reveal the subnuclear compartments containing the DNA. Moreover, it was revealed that there exist small particulate structures outside the DNA-containing territories and they were redefined with the arrival of different autoantigen antibodies in the 1990s. The recent advances in light microscopy and the use of live cell fluorophores revealed to (and mostly surprised) this generation of cell biologists that the components within the subnuclear structures (Chen and Huang, 2001; Misteli, 2001; Reits and Neefjes, 2001), and sometimes these structures themselves, are not static but rather are dynamic (Boudonck et al., 1999; Platani et al., 2000; Snaar et al., 2000; Muratani et al., 2001; Platani et al., 2002). At the same time, biochemical procedures have been developed to isolate different nuclear subcompartments, in particular the nucleolus, for functional studies (Section I.3.1). So the underlying questions are (1) why is the nucleus subcompartmentalised? and (2) what are the respective functions of these subcompartments? This thesis is aimed at understanding more about the functions and dynamics of the human nucleolus, the most prominent subnuclear compartment.

Despite over 150 years of research, the nucleolus was first described as being a ribosome factory only 40 years ago; yet even today we know little about the components of the ribosome production machineries in human (Fatica and Tollervey, 2002). Moreover, the idea of the nucleolus being a ribosome factory cannot explain the presence of viral, cell cycle regulatory and tumor-related proteins within this structure (Section I.2.5). This chapter aims at giving an overview of the properties of different subnuclear structures (Section I.1) and particularly will focus on the current status of our understanding of the nucleolus (Section I.2). This is followed by an introduction of the analytic tools (Section I.3) that were used during this PhD (Section I.4).

I.1.2 Compartmentalisation of the Nucleus

A compartment is mathematically defined as a quantity of a substance that has uniform and distinguishable kinetics of transport (Godfrey, 1983). Within the cell context, a compartment is often defined by a physical separation – a membrane – for example, the cytoplasmic organelles such as the mitochondria and the endoplasmic reticulum (ER). However, a membrane is not necessary for compartmentalisation provided that the contents within are homogenous, well-mixed and distinct from the surroundings. The mammalian nucleus is highly organised itself into different compartments (or subnuclear structures) that, even without intra-nuclear membranes, contain distinct set of proteins and are separate from one another (Figure I-1).

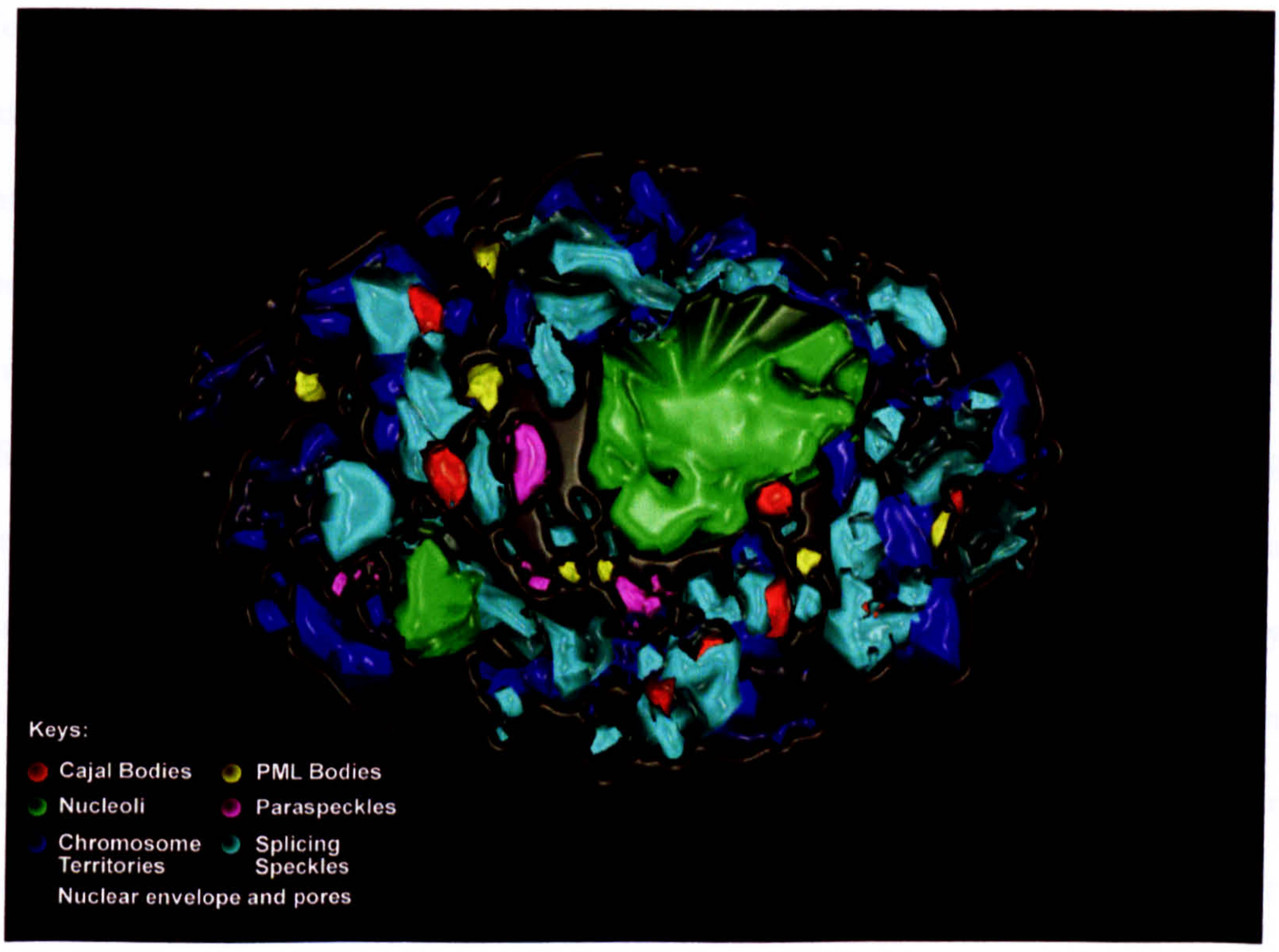


Figure I-1 **Model of Nuclear Architecture.** This model illustrates two properties of the nucleus: (a) it is very compacted and dense; and (b) it is non-uniform and compartmentalized. Red: Cajal bodies; Green: Nucleoli; Blue: Chromosome territories; Yellow: PML bodies; Magenta: Paraspeckles and Cyan: Splicing speckles and the nuclear envelope and pores are denoted by a dashed circle.

Each chromosome, and even DNA sequences with differential timing of replication within the chromosomes, occupies a nuclear space or territory (Schul et al., 1998; Boyle et al., 2001; Cremer and Cremer, 2001; Dietzel and Belmont, 2001; Mahy et al., 2002a; Mahy et al., 2002b; Parada and Misteli, 2002). Within the intrachromosomal spaces, a different set of proteins were identified in specific patterns (Lamond and Earnshaw, 1998; Matera, 1999; Dundr and Misteli, 2001; Spector, 2001; Swedlow and Lamond, 2001; Carmo-Fonseca, 2002; Carmo-Fonseca et al., 2002); some are associated with chromosomes such as the nucleolus and OPT domain (Table I-1) and some are not such as PML bodies and splicing speckles; some structures such as Cajal bodies and PML bodies are continually moving within the nucleoplasm whilst others such as splicing speckles are relatively immobile in terms of their nuclear positions (Matera, 1999; Platani et al., 2000; Snaar et al., 2000; Muratani et al., 2001; Swedlow and Lamond, 2001; Carmo-Fonseca et al., 2002; Ogg and Lamond, 2002; Platani et al., 2002). However, whether the subnuclear structure itself is static or dynamic, associated with chromosomes or not, it seems that they share two interesting properties: (1) the subnuclear structures are disassembled when the nuclear envelope breaks down during mitosis and they are reassembled when the nuclear envelope reforms, and (2) most of the components studied so far are continuously exchanging between the nucleoplasm and the subnuclear structure, though each is comprised of a distinct sets of proteins. It is however beyond the scope of this thesis to discuss every subnuclear structure and I will only focus on a few structures that are related to the nucleolus (Section I.1.3) and the nucleolus itself (Section I.2). The properties of subnuclear structures are summarised in Table I-1 and recently reviewed elsewhere in detail (Singer and Green, 1997; Lamond and Earnshaw, 1998; Matera, 1999; Cremer and Cremer, 2001; Dundr and Misteli, 2001; Gasser, 2001; Spector, 2001; Swedlow and Lamond, 2001; Wolffe and Hansen, 2001; Carmo-Fonseca, 2002; Carmo-Fonseca et al., 2002; Iborra and Cook, 2002; Pederson, 2002).

Subnuclear structures	Properties
Cajal Body (section I.1.3.1)	Ribonucleoprotein particle (RNP) maturation; histone gene transcription?
Chromosome territories	Chromosomes occupy distinct nuclear space relative to the nuclear periphery
Gems	Colocalised with Cajal body in most cell types; containing marker protein SMN
Nucleolus (section I.2)	Ribosome factories; cell cycle control; viral target
OPT domain	Foci containing Oct1 and PTF transcription factors; often associated with human chromosomes 6 and 7
Paraspeckles (section I.1.3.2)	Transcriptional inhibition-dependent association with the nucleolus; often located next to splicing speckles; containing paraspeckle protein 1 (PSP1)
Perinucleolar compartment (PNC; section I.1.3.2)	Rarely found in primary cells or non-cancerous cells; associated with the nucleolar periphery; containing polypyrimidine tract binding protein (PTB) and RNA polymerase III transcripts
PcG domain	Containing *polycomb* group proteins; probably involved in gene repression
PML Body	Anti-viral response; enriched for sumolyated protein; containing marker protein PML, sp100 and CREB-binding protein (CBP)
SAM68 bodies (section I.1.3.2)	Associated with the nucleolar periphery; containing proteins RNA-binding proteins with the STAR (signal transduction and activation of RNA) domain; commonly observed in cancerous cells.
Splicing speckles	Storage site of splicing factors? Recycling site of splicing factors?

Table I-1 **Subnuclear structures in the interphase nucleus.**

I.1.3 Cajal Bodies and other Nucleolus-related nuclear structures

I.1.3.1 Cajal Bodies

Cajal bodies, originally called nucleolar accessory bodies, or coiled bodies, were recently renamed after their discoverer, Ramón y Cajal (Cajal, 1903; Gall, 2000). They are usually defined by the presence of the marker protein, p80 coilin, and are often associated with the nucleolar periphery, or even located inside the nucleolus (Malatesta et al., 1994; Ochs et al., 1994; Lyon et al., 1997; Sleeman et al., 1998). Typically there are 1-10 Cajal bodies ranging from 0.2 to 1.0 μm or larger in size within a human cell (Lamond and Earnshaw, 1998; Platani et al., 2000). They contain small nuclear ribonucleoprotein particles (snRNPs), some transcription factors and several nucleolar proteins, including fibrillarin, a component of

small nucleolar ribonucleoprotein particles, or snoRNPs, (Gall, 2000). However, they do not contain non-snRNP protein splicing factors or nascent pre-mRNA. Their snRNP components are specifically only the newly assembled particles (Carvalho et al., 1999; Gall et al., 1999; Sleeman and Lamond, 1999). Fluorescently labeled U3, U8 and U14 snoRNAs, when microinjected into *Xenopus* oocytes, accumulate transiently in Cajal bodies prior to nucleoli, suggesting that newly imported snoRNAs flow from the Cajal bodies to nucleoli (Narayanan et al., 1999a; Narayanan et al., 1999b). Therefore, Cajal bodies might be involved in transport and maturation of both snRNPs and snoRNPs.

In addition to small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs), a new class of RNAs which has two different types of classical snoRNA motifs (box C/D and H/ACA; section I.2.3) was recently found to be exclusively localised in Cajal bodies, and was therefore named small Cajal body-specific RNA (scaRNAs) (Darzacq et al., 2002; Kiss et al., 2002). Analogous to the function of many snoRNAs which modify certain regions of ribosomal RNA (rRNA), scaRNAs contain a region that can potentially be base-paired with a complementary region of snRNA and hence may be involved in 2'-O-methylation and pseudouridylation of the spliceosomal snRNAs U1, U2, U4 andU5 within Cajal bodies.

An important aspect of subnuclear structures is their dynamic and regulated association with one another. This is especially evident for the nucleolus. Even at the light microscopy level back in the early 20th century, a close association between Cajal bodies and nucleoli had been observed (Cajal, 1903; Gall, 2000). Recent studies suggest that a number of Cajal body components are shared with the nucleolus, particularly with the dense fibrillar component (Section I.2.4). Dynamic tracking of Cajal bodies in plant and human cells showed that Cajal bodies can enter into nucleoli (Boudonck et al., 1999; Platani et al., 2000; Snaar et al., 2000).

I.1.3.2 Paraspeckles, Perinucleolar compartments and SAM68 bodies

During the identification of the human nucleolar proteome, our group has identified a new subnuclear structure known as **paraspeckles** that are associated with the nucleous in a transcription inhibition-dependent manner (Fox et al., 2002). The marker protein paraspeckle protein 1 (PSP1) is continuously exchanging between the nucleolus and paraspeckles and PSP1 relocates to the nucleolar periphery upon transcriptional inhibition. In contrast, two structures which normally reside in the nucleolar periphery, namely the **perinucleolar compartment** (PNC) and the **SAM68** nuclear body, are disassembled upon transcriptional inhibition (Huang, 2000). PNC is a subcompartment associated with nucleoli in cancer cell, but not in primary cells, that contains the marker protein polypyrimidine tract binding (PTB) proteins and polymerase III transcripts (Huang et al., 1997; Huang, 2000). The presence of PNCs and the size of nucleoli have been used as a prognostic indicator of cell malignancy. Why do these subnuclear structures (or their individual components) associate with/depart from the nucleolus under different metabolic conditions? Which nucleolar components attract or repel them? The underlying basic question is, "what constitutes the nucleolus?"

I.2 Overview of the Nucleolus

I.2.1 Historical Perspectives

Between 1835 and 1838, the nucleolus was discovered in different systems concurrently by Wagner, Valentin and Schleiden (Franke, 1988). Gabriel Gustav Valentin is the first person who described the nucleolus in somatic cells as stated in the quote in the beginning of this chapter, which translates as: '*In every cell, without exception, there exists a somewhat darker-appearing and compact nucleus of a round or nearly round shape. Mostly it is located in the centre of the specific cell, composed of a finely granular material and containing in its interior an exactly spheroidal body which in this way forms a kind of secondary nucleus within the first one*' (Valentin, 1836). Since the nucleus is

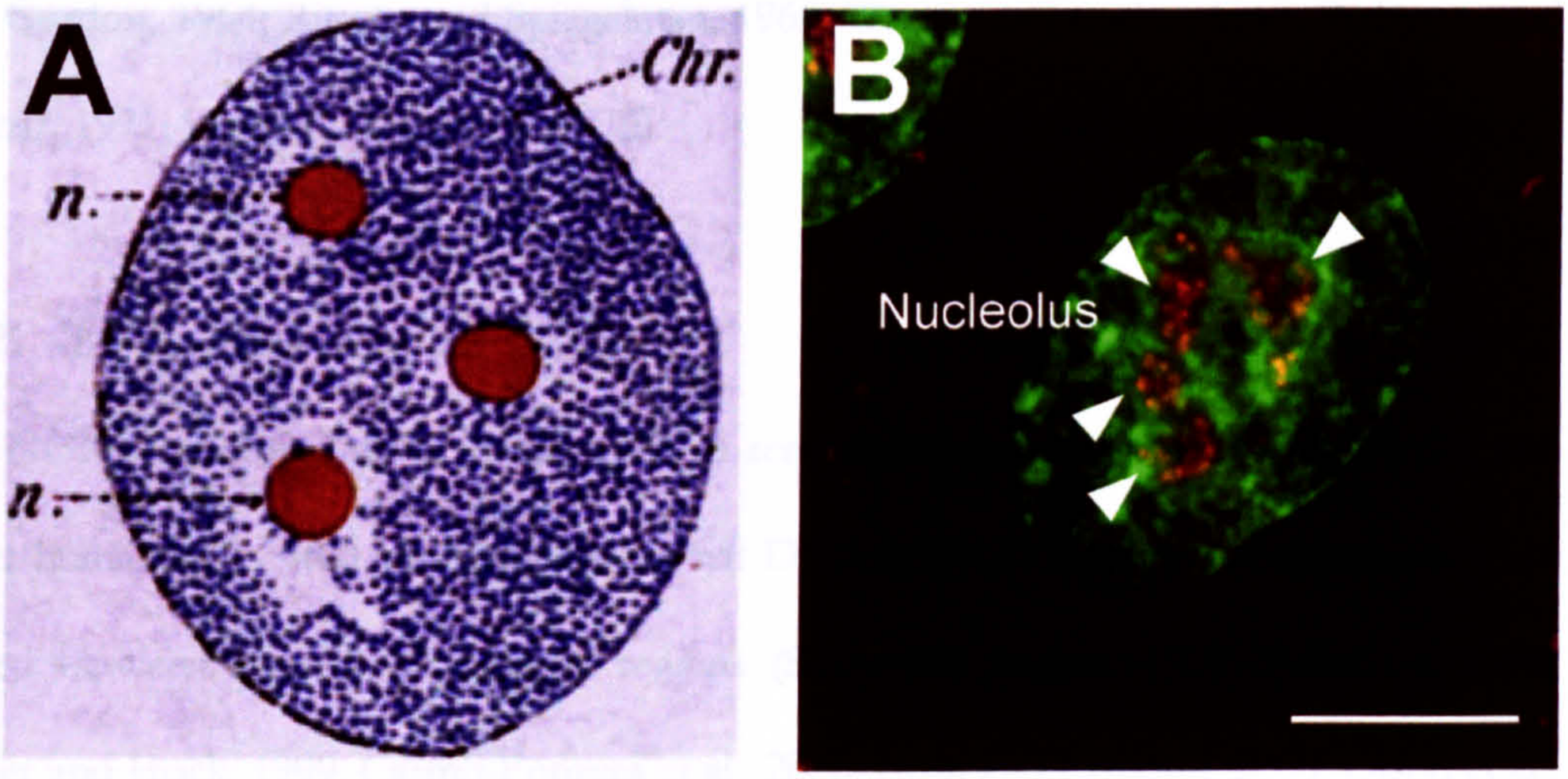


Figure I-2 **Nucleolus – 'a secondary nucleus within the nucleus'.** (A) An early drawing taken from Montgomery 1898, illustrating the location of nucleoli inside the nucleus and (B) A light micrograph taken using a DeltaVision Restoration microscope, illustrating nucleoli and the DNA content within the nucleus from a HeLa cell line expressing GFP-H2B (green) which was stained with anti-fibrillarin antibody 72B9 (red). Scale bar =5μm. Courtesy of Prof. Thoru Pederson for sharing the early publication regarding the nucleolus.

dense and highly refractile, it was the first subnuclear organelle studied by microscopists (Figure I-2). Both the variability in number and size and the disappearance and reappearance of this organelle during mitosis were noted as early as the late 19th century (Montgomery, 1898).

In the 1930s, Heitz and subsequently McClintock have observed that the number and lengths of less compact regions or secondary constrictions in mitotic chromosomes are related to the number and sizes of nucleoli and these constrictions are now known as nucleolar organizing regions or NORs (Heitz, 1931; McClintock, 1934). However, the function of NORs was not clear until the 1960s when it was shown that the NOR contains genes encoding for ribosomal 18S, 5.8S and 28S RNAs and mutants that lack NORs are devoid of nucleoli (Ritossa and Spiegelman, 1965; Birnstiel and Chipchase, 1970). Soon thereafter the presence of ribosomal proteins and the assembly of nascent ribosomes in the nucleolus were also revealed, establishing that the nucleolus is the site of ribosomal RNA synthesis and ribosome assembly, i.e. a ribosome factory (Birnstiel and Hyde, 1963; Brown

and Gurdon, 1964; Ritossa and Spiegelman, 1965; Warner and Soeiro, 1967; Pederson and Kumar, 1971; Kumar and Warner, 1972).

I.2.2 The Nucleolus: its DNA

The NORs are located on the short arms of acrocentric chromosomes 13, 14, 15, 21, and 22 in human and ~400 copies of ribosomal DNA (rDNA) repeats in tandem array are shared between these chromosomal regions (Shaw and Jordan, 1995; Pederson, 1998; Scheer and Hock, 1999; Carmo-Fonseca et al., 2000; Lyon and Lamond, 2000; Olson et al., 2000; Pederson and Politz, 2000; Visintin and Amon, 2000). Each repeat of about 43 kbp contains the transcribed regions of 18S, 5.8S and 28S rRNA genes plus a non-transcribed spacer. The high number of repeats renders the rDNA loci difficult to analyse by sequencing from end to end (Labella and Schlessinger, 1989), and it is generally believed, but not proven, that the rDNA repeats are arranged as head-to-tail clusters in a telomere-to-centromere orientation. The gene that encodes the remaining ribosomal RNA, i.e. 5S rRNA is usually located outside the nucleolus in higher plant and animal cells, but the transcript itself is, surprisingly, localised in the nucleolus (Section I.2.3).

However, even in metabolically active cells, only a subset of rDNA copies is transcribed and at most only ~50% of the copies are utilized (Moss and Stefanovsky, 2002). The ratio of active and inactive rRNA genes is tissue specific and is stably propagated through the cell cycle. Recent studies identified a nucleolar remodeling complex (NoRC) that mediates heterochromatin formation and silencing of ribosomal gene transcription by recruiting methyltransferase and histone deacetylase activities to the rDNA promoter (Strohner et al., 2001; Santoro et al., 2002; Zhou et al., 2002a). It is hypothesised that this constitutive silencing of a significant fraction of the tandemly repeated rDNA genes is required to protect the genes against unwanted homologous recombination.

The rDNA organization as tandem repeats in heterochromatic regions is conserved amongst eukaryotes. However, these conserved features are not required for the proper function of rDNA. For example, a single rDNA gene inserted within a euchromatin region in the Drosophila genome is active in transcription and results in formation of a nucleolus at the integrated site (Karpen et al., 1988). Similar results were obtained in human and higher plants provided that the promoter regions were included (Hadjiargyrou et al., 1994; Wanzenbock et al., 1997). Transcription of plasmid containing a single rDNA locus driven by an RNA polymerase II promoter in a yeast mutant devoid of rDNA repeats, resulted in formation of multiple, mini-nucleoli throughout the nucleoplasm (Nomura, 1999). Conversely, if the plasmid was specifically designed to be transcribed by RNA polymerase I, a single large nucleolus formed. This raises an interesting issue of whether active transcription or a specific sequence in the repeat itself is responsible for the recruitment of rDNA and coalescence into nucleoli. In yeast, rDNAs are specifically associated with the Net1 protein that can recruit a member of the Silent Information Regulatory (SIR) family, Sir2 (Straight et al., 1999; Carmo-Fonseca et al., 2000). SIR proteins are known for their association with the other family members and histones to orchestrate silencing in yeast heterochromatic regions, such as the silent mating-type (HM) loci and telomeres. Similarly, rDNAs in human or other higher organisms maintain a tight association with heterochromatin and this compartment contains DNA sequences from several chromosomes, including those that are devoid of NORs (Pluta et al., 1995; Carmo-Fonseca, 2002). Interestingly, it was shown that even transcriptionally silent human ribosomal genes within a murine cell environment are incorporated into nucleoli *in vivo* (Sullivan et al., 2001). This led to the suggestion that the clustering of rDNA may result from protein-protein interactions between nearby heterochromatin domains located in either the same, or distinct, chromosomes.

I.2.3 The Nucleolus: its RNA

Another amazing aspect of the nucleolus is the plethora of RNA species which localise there, regardless of whether the RNA is transcribed there or not (Maxwell and Fournier, 1995; Lafontaine and Tollervey, 1998; Weinstein and Steitz, 1999). An actively cycling eukaryotic cell expends ~35-60% of its total transcriptional activity in transcribing the rDNA repeats and another 10-20% for transcribing 5S rDNA, ~150 different species of snoRNAs and other RNAs such as tRNAs and signal recognition particle (SRP) RNAs. This constitutes ~50% and ~80% of the total transcriptional activity in mammal and yeast cells, respectively (Warner, 1999).

In eukaryotes, RNA polymerase I is responsible for the rDNA transcription and is localised in a specific subnucleolar region known as the fibrillar centre (Section I.2.4). Transcription requires the binding of transcription factors UBF and SL1 and these factors are phosphorylated and repressed by cyclin B-cdk1 kinase during mitosis (Klein and Grummt, 1999; Voit et al., 1999; Sirri et al., 2000; Voit and Grummt, 2001; Hernandez-Verdun et al., 2002; Sirri et al., 2002). Surprisingly, both the polymerase itself and other accessory transcription factors are localised to mitotic chromosomes, notably at the NORs, and transcription at these sites can be artificially induced during mitosis by cdc2 kinase inhibitor (Scheer et al., 1984; Scheer and Rose, 1984; Roussel et al., 1993; Zatsepina et al., 1993; Gilbert et al., 1995; Weisenberger and Scheer, 1995; Jordan et al., 1996; Gebrane-Younes et al., 1997; Sirri et al., 2000). During interphase, the rate of rDNA transcription is regulated in accordance with cell growth and the need for new ribosomes (Carmo-Fonseca et al., 1996; Warner, 1999; Moss and Stefanovsky, 2002). When cells approach a stationary phase or are serum-starved, rDNA transcription is down-regulated. There seems to be a direct relationship between nucleolar size and the transcriptional state of the cell (Stefanovsky et al., 2001).

Active ribosomal RNA genes can be visualized in EM by a simple hypotonic spreading technique, now known as "Miller spread", after the name of one of the investigators (Miller and Beatty, 1969). It was shown that the length of nascent transcripts are increasingly branching along the axis in one direction, with many engaged RNA polymerases on the transcribed regions of rDNA, resulting in a 'Christmas tree pattern', whilst the spacer regions remain empty. Moreover, knob-like structures exist at the termini of the nascent transcripts and these structures contain proteins required for rRNA processing (Dragon et al., 2002).

The rRNA processing involves cleavage and modification steps to generate mature rRNA species that form the catalytic core of the ribosome (Warner, 2001). Various snoRNAs (http://ma.wustl.edu/snoRNAdb/), for example U3 and U14, are needed for these separate rRNA cleavage steps, each snoRNA being required at specific steps. However, the exact nature of the cleavage sites in human is rather poorly understood, compared with yeast and mouse. It is also not clear whether rRNA processing requires the same level of precision as pre-mRNA splicing because there is no reading frame to maintain (Bowman et al., 1983; Eichler and Craig, 1994). Limited heterogeneity of a few nucleotides has been reported for the ends of 5.8S and 28S rRNA in functioning ribosomes; therefore the cleavages and trimming may be rather approximate. Similar to pre-mRNA splicing, the existence of alternative rRNA processing pathways is evident in different mammalian tissues (Akhmanova et al., 2000).

The snoRNAs can be categorized into two main classes – box C/D (e.g. U3, U8, U14) and box H/ACA (e.g. U17, U19, U64) – according to conserved sequence elements and the way they are assumed to fold into defined secondary structures (Figure I-3) (Weinstein and Steitz, 1999). These snoRNAs assemble together with protein factors to form snoRNPs

that have roles in guiding and catalyzing post-transcriptional RNA modifications. For example, box C/D snoRNPs direct 2'-O-ribose methylation and box H/ACA snoRNPs direct pseudouridylations of specific rRNA nucleotides (Lafontaine and Tollervey, 1998; Lewis and Tollervey, 2000). Interestingly, the modifications themselves are not all essential for cell growth, because a yeast strain lacking as many as eight different snoRNAs can still grow normally; however, the modifications are conserved and serve to enhance the conformational stability of the rRNA and hence the functional efficiency of the ribosome. Another interesting aspect of snoRNAs in vertebrates is that they are housed within the introns of pre-mRNAs that share the 5'- pyrimidine-rich terminus characteristic of the 5' Terminal OligoPyrimidine (5'TOP) family. Members of the 5' TOP family include ribosomal proteins and many other nucleolar proteins, such as the chaperone B23 (Maxwell and Fournier, 1995; Lafontaine and Tollervey, 1998; Weinstein and Steitz, 1999).

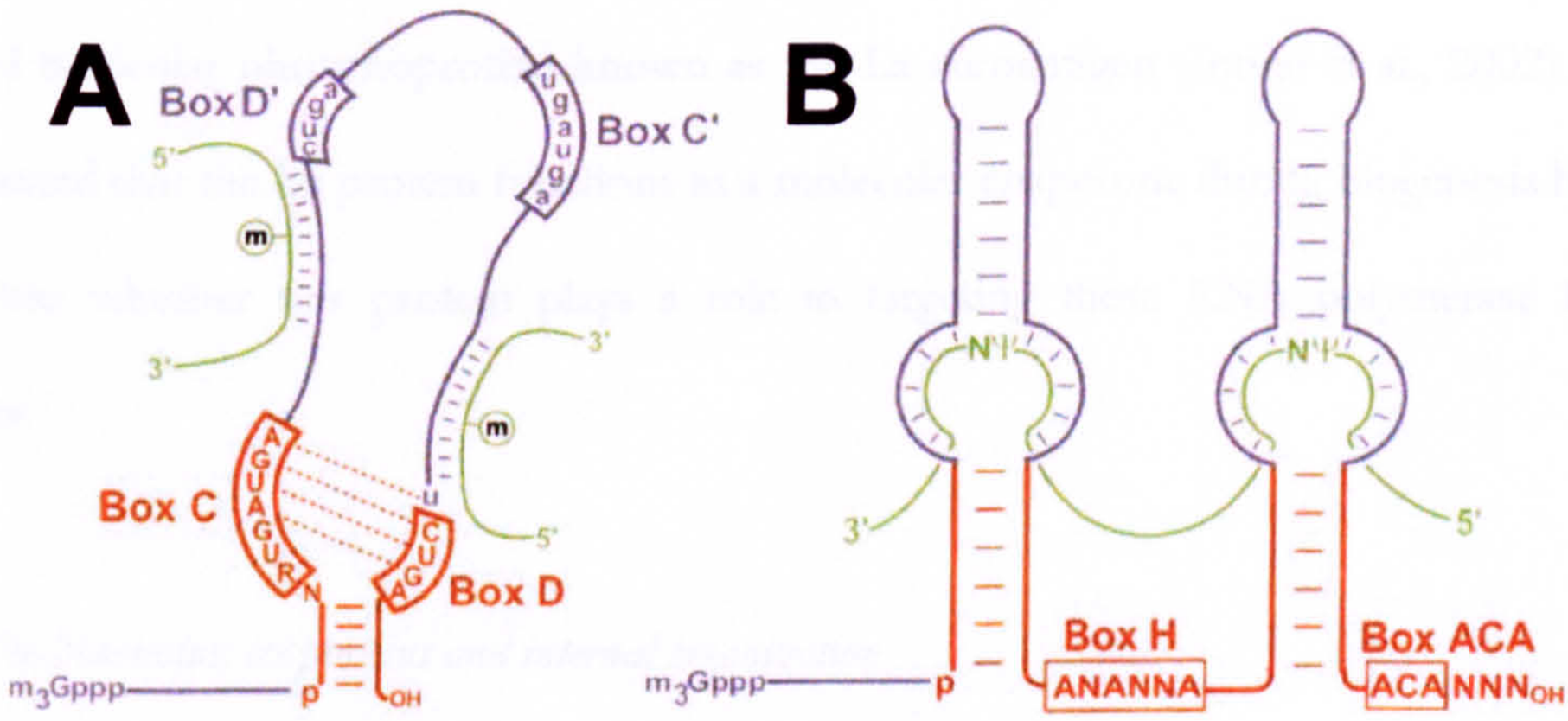


Figure I-3 **Conserved sequence and structural features of snoRNAs.** (A) Box C/D snoRNAs form 10-21nt duplexes with the substrate rRNA. The methyl group is added to the rRNA nucleotide paired to the fifth residue upstream of box D and/or D'. Boxes D' and C' are typically 3-9nt apart. (B) Box H/ACA snoRNAs have a conserved hairpin-hinge-hairpin-tail secondary structure. Each hairpin forms two short duplexes (4-10nt) with the rRNA, leaving the target uridine unpaired in the 'pseudouridylation pocket Ψ' N, any nucleotide. Adapted from Kiss (2002).

Apart from all the ribosomal RNAs and snoRNAs, it was found that SRP RNA is also localised in nucleoli as are the SRP proteins SRP19, SRP68 and SRP72, but not ER-bound SRP54, which together they form a complex to direct proteins into ER for secretion (Pederson, 1999; Politz et al., 2000; Politz et al., 2002). Similarly, RNase P, the ribonucleoprotein enzyme that mediates the 5' processing of pre-tRNAs, and some tRNAs are present in nucleoli (Jarrous et al., 1999; Jarrous et al., 2001). However, the roles that these classes of RNAs play in the nucleolus remain to be established, although it was speculated that these factors, which are required for translation in the cytoplasm, may be exported together with the ribosomal subunits in a preassembled state. The spliceosomal U6 snRNA has also been shown to enter the nucleolus where it is modified by snoRNAs (Lange and Gerbi, 2000; Lewis and Tollervey, 2000). A common feature shared by the 5S rRNA, tRNA, RNase P RNA, SRP RNA and U6 snRNA is that all RNAs are transcribed outside the nucleolus by RNA polymerase III (Carmo-Fonseca et al., 2000; Lewis and Tollervey, 2000). All newly synthesized polymerase III transcripts bind to a highly conserved nucleolar phosphoprotein known as the La autoantigen (Intine et al., 2002). It was suggested that the La protein functions as a molecular chaperone during biogenesis but it is unclear whether this protein plays a role in targeting these RNA polymerase III transcripts.

I.2.4 The Nucleolus: its proteins and internal organization

In chromosome preparations, the NOR can be very clearly and rapidly visualized by ammoniacal or formic acid silver nitrate staining (Derenzini et al., 1998). The silver stainability of the NOR is due to the presence of a peculiar set of acidic proteins, which are highly argyrophilic. This led to the initial identification of a handful of human nucleolar autoantigens, including RNA polymerase I transcription factor UBF, the molecular chaperone for ribosomal assembly, B23, and the highly abundant and highly

phosphorylated protein C23, or nucleolin. Nucleolin/C23 alone constitutes ~10% of total nucleolar proteins by mass. An exhaustive list of known nucleolar proteins was compiled in Chapter III.

With the availability of the antibodies against nucleolar proteins, it was soon realized that these components localised specifically to morphologically distinct regions within the nucleolus as revealed by conventional thin section electron microscopy (Figure I-4). These regions are the fibrillar centre (FC), the dense fibrillar component (DFC) and the granular component (GC). FCs are pale fibrillar regions that are enriched for RNA polymerase I, transcription factors involved in rDNA transcription, such as UBF and DNA topoisomerase I; DFCs are highly electron dense fibrillar regions that partially, or completely, surround FCs and are heavily labeled with antibodies that recognize pre-rRNA processing factors, such as fibrillarin; GC, the granular component, that is enriched for

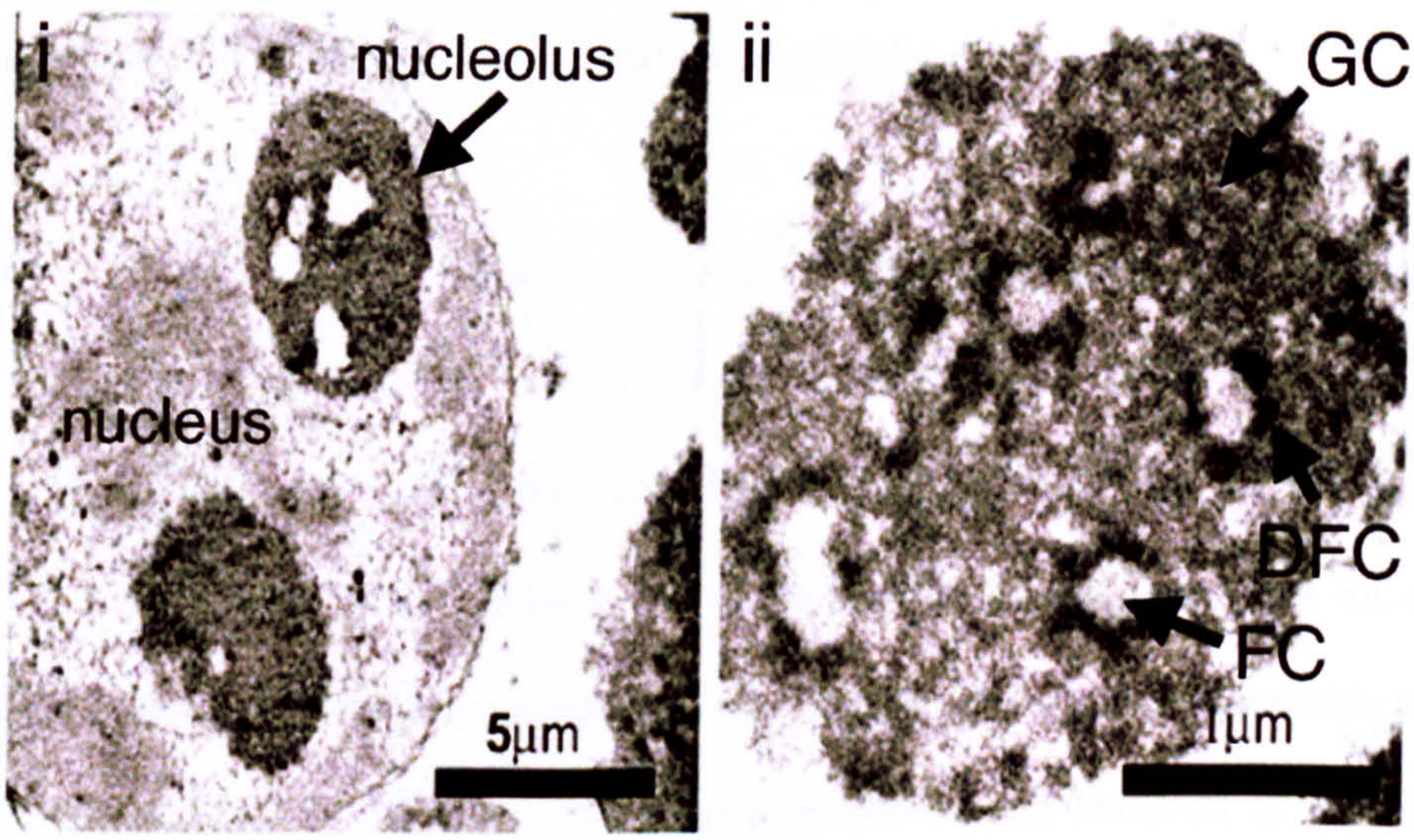


Figure I-4 **Internal organization of the Nucleolus.** Transmission electron micrograph of thin sections stained by uranium and lead salts showing the internal morphology of the nucleolus *in situ* on panel i; scale bar = 5µm; and the close up of different domains on panel ii; FC—fibrillar centre; DFC—dense fibrillar component; GC—granular component; scale bar = 1 µm. Picture courtesy of Carol Lyon.

ribosomal proteins and ribosome subunit assembly factors, constitutes the rest of the nucleolus surrounding the DFCs and the FCs, (Shaw and Jordan, 1995; Scheer and Hock, 1999; Carmo-Fonseca et al., 2000).

The location of the active rDNA genes, i.e. the 'Christmas tree' pattern *in vivo*, within the nucleolus, however, remains controversial. Early autoradiographic work at the EM level showed that the DFCs were the likely site of RNA synthesis and the granular component a site of late processing and pre-ribosome accumulation (Granboulan and Granboulan, 1965). Subsequent pulse-chase experiments using the halogenated analogue of UTP, BrUTP, have suggested a mixed scenario. Depending on the protocols employed by different laboratories, the site of incorporation could be either at the FC, or FC/DFC borders or DFCs (Hozak et al., 1994; Thiry et al., 2000; Cheutin et al., 2002; Huang, 2002; Koberna et al., 2002). Nevertheless, it is obvious from all these experiments that the pre-rRNAs migrate from fibrillar regions to granular regions vectorially and subsequently into the nucleoplasm while maturing into preribosomal particles.

The number of FCs is directly related to the rate of rRNA synthesis (Figure I-5A-B). For example, it may fall from ~234 to ~156 upon serum starvation of fibroblasts or rise from ~9 to ~80 upon stimulation of a peripheral blood lymphocyte (Jordan and McGovern, 1981; Haaf et al., 1991). This suggests that each fibrillar centre represents an active transcription factory, and when transcription is upregulated, new factories assemble on the once silent rDNA repeat. However, the molecular mechanism involved in the rDNA switching from a repressed to an activated state remains to be established (Section I.2.2).

Akin to other subnuclear structures, the nucleolus is very dynamic, showing great variability as a function of activity in animal and plant cells (Section I.2.3; Figure I-5C- D). Fully active

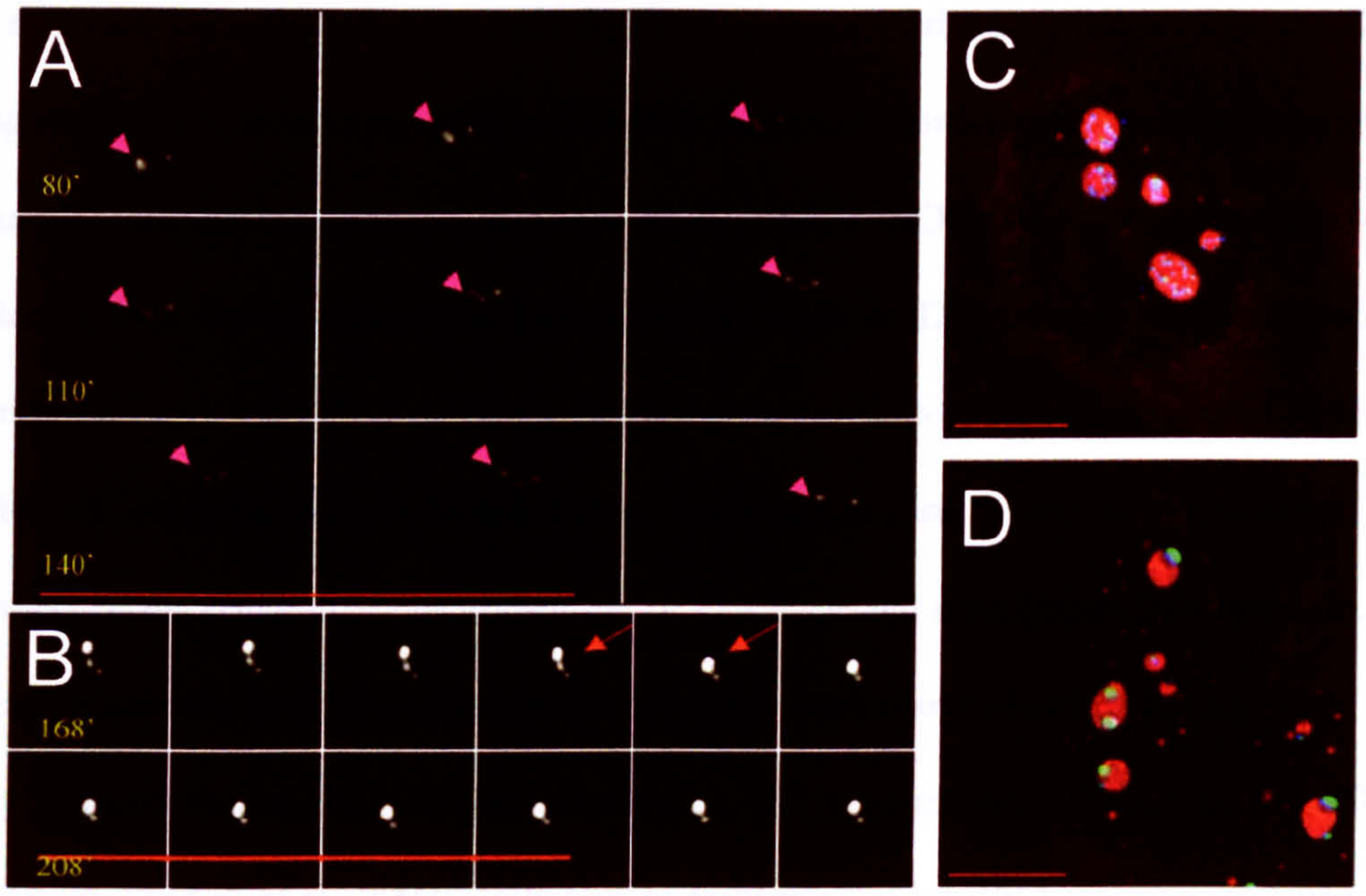


Figure I-5 **Dynamic organization of Subnucleolar Domains.** Live cell analyses of HeLa cells stably expressing EYFP-RNA polymerase I subunit RPA39, which is localised in FCs, (A) upon release from serum starvation and (B) upon transcriptional inhibition (0.04μg/ml actinomycin D). Series showing (A) from 80-160min after release and (B)168-247min after inhibition; scale bar = 20μm. (A) Purple arrowheads indicate a splitting event of FCs and (B) red arrows indicate a merging event of FCs. HeLa cells stably expressing EYFP-RPA39 were fixed (C) before and (D) after 3 hour of actinomycin D treatment (0.04μg/ml) and stained with anti-fibrillarin antibodies 72B9 for DFC, and pyronin Y, a non-specific RNA dye, which stains the cytoplasm and GCs in the nucleolus; scale bar=5μm.. Green: FC (EYFP-RPA39); Blue: DFC (fibrillarin) and Red: GC (pyronin Y).

nucleoli are large, with extensive intermingling of the FCs, the DFCs and the GCs, resulting in a reticulate-like structure (also termed nucleolonema). In contrast, in either some quiescent cells, or those with nucleolar inactivation following induced transcriptional arrest, the components of the FCs, the DFCs and the GCs are segregated out into adjacent blocks.

For cells that undergo a 'closed mitosis', such as budding yeast, the nucleolus remains intact during mitosis, ultimately being separated along the mitotic spindle during the late stage of division (Carmo-Fonseca et al., 2000). In contrast, the nucleolus from multicellular organisms disassembles during mitosis (Shaw and Jordan, 1995). It was previously shown

that nucleolar breakdown begins with the disappearance of the GC, followed by the DFC in higher plants (Lafontaine, 1968) and various nucleolar proteins leave the nucleolus in an apparently ordered progression (Gautier et al., 1992). NORs are ultrastructurally very similar to FCs during mitosis and components such as DNA topoisomerase I, RNA polymerase I and other RNA polymerase I transcription factors remain associated with the condensed chromosomes, whereas other components are found either around the periphery of the condensed chromosomes, or else dispersed throughout the cytoplasm. At the end of mitosis, small round structures called pre-nucleolar bodies (PNBs) appear in the newly formed nuclei, before the resumption of transcription (Dundr et al., 2000; Hernandez-Verdun et al., 2002). Initiation of transcription by RNA polymerase I triggers the recruitment of PNBs to the NORs to form a complete nucleolus. Taken together, these data indicate that the assembly of a nucleolus requires the rRNA transcripts.

Many proteins, such as ribosomal proteins, translocate once from the cytoplasm to the nucleolus and then travel back to the cytoplasm as pre-ribosomal particles (Section I.2.5.1). Other nucleolar proteins surprisingly shuttle back and forth continually between the nucleolus and the cytoplasm. Intensive investigation has been carried out to identify a hypothetical nucleolar targeting signal. However, most of the sequences that have been found seem to function as nuclear targeting signals rather than nucleolar targeting signals (Hatanaka, 1990). Some nucleolar targeting signals identified, for example within the human retrovirus HIV-1 Rev and Tat proteins, seem to bind to endogenous nucleolar proteins such as the molecular chaperone B23, for nucleolar entry (Hiscox, 2002). Recent fluorescence recovery after photobleaching (FRAP) analyses have indicated that proteins related to ribosome biogenesis, including RNA polymerase I subunits, transcription factors, factors required for rRNA processing and ribosome assembly are rapidly shuttling between the nucleoplasm and the nucleolus (see Chapter VI). However, if that is the case, this raises

an interesting question as to whether the signal required for the apparent nucleolar accumulation is targeting proteins to the nucleolus either via 'receptor' recognition, or by changing the rate of molecular interaction with other partners within the nucleolus. If the latter case is correct, as suggested by the FRAP analyses, then what are the molecular entities that attract protein to accumulate in the nucleolus? And why are those entities retained there in the first place? These two basic questions deserve to be answered, not only with regard to the nucleolus, but also for all other subnuclear structures.

I.2.5 Proposed roles of the Nucleolus

I.2.5.1 The Nucleolus as a ribosome factory

As discussed in previous sections (Sections I.2.1-I.2.4), both rDNA transcription and subsequently rRNA processing and modification occur within the nucleolus, which is followed by ribosome subunit assembly. Over the last decades, insight into ribosomal subunit assembly has mainly come from studies on budding yeast, especially after the introduction of a GFP-based ribosome export assay (Hurt et al., 1999). Currently, the export mechanism of the pre-40S particles (the small ribosomal subunit) is still poorly understood compared with the pre-60S particles (the large ribosomal subunit) (Rout et al., 1997). Recent large-scale systematic purification of protein-complexes using high affinity tags has led to tremendous progress on how the pre-60S particles are formed. This is even more impressive given that the half-lives of these transient structures are only about 1 minute *in vivo* (Aitchison and Rout, 2000; Ho et al., 2000; Gadal et al., 2001b; Nissan et al., 2002). To better define the pathway of 60S subunit synthesis, a series of particles was pulled out using tagged version of several previously identified components from the pre-60S particles. These studies revealed a series of 'snapshots' of the pre-60S ribosomes as they move from the nucleolus to the cytoplasm.

In these studies, a member of the 'AAA' family of ATPases – ATPase associated with various cellular activities – Rix7 was identified (Gadal et al., 2001a). The ATPase family is involved in many cellular functions, but in all cases they occur in situations where protein unfolding or dissociation occurs, e.g. at microtubules (Vale, 2000). It is currently proposed that Rix7 facilitates changes in either protein-RNA, or protein-protein, interactions within the pre-60S particles. As the complexes exit the nucleolus, Rix7 facilitates the shedding of processing factors during transit to the nuclear periphery (Brown, 2001). This allows the formation of an export-competent pre-60S particle, which is then complexed with a pre-ribosomal particle-specific adapter of an export receptor, Nmd3. The complex is then exported via the nuclear pore to the cytoplasm by a Ran-dependent nuclear export receptor, XpoI. Recently, Politz et al used fluorescently labelled oligos that specifically hybridise to non-conserved regions of the 28S rRNA to show that large ribosomal subunits move out randomly from the nucleolus and into the nucleoplasm in all directions in human cells, with no evidence of concentrated movement along a directed path (Politz J, personal communication). Much remains to be done to solve the puzzle of ribosome biogenesis and export in yeast and, even more, in human. Nevertheless, the foundation laid in yeast and the availability of the soon-completed human genome may help speed up our understanding of these processes.

I.2.5.2 The Nucleolus as a molecular 'safe' or 'sink'

Data obtained in the past few years suggest that the nucleolus may play a role in cell cycle regulation by sequestration of proteins. The tumor-suppressor protein p53 is a transcription factor that switches on a series of protective genes when the cell is exposed to stress (Lohrum et al., 2000; Sherr and Weber, 2000). Under normal conditions, p53 activates the MDM2 gene for expression. MDM2 mediates the ubiquitination of p53 and hence its degradation in the cytoplasm. Both MDM2 and p53 shuttle between the nucleus

and the cytoplasm, but MDM2 contains a cryptic nucleolar signal that does not function in unstressed cells. This signal is necessary to cooperate with the nucleolar tumour suppressor ARF to allow the relocalisation of both proteins into the nucleolus, thereby releasing p53 from the MDM2-degradation pathway to activate the stress response pathway.

Recent studies, tagging GFP to the telomerase protein component TERT, revealed that telomerase activity is related to subnuclear localisation (Wong et al., 2002; Yang et al., 2002). The release of sequestered telomerase to the nucleoplasm is enhanced at the expected time of telomere replication, i.e. S-phase. By contrast, in tumor and transformed cells, there is an almost complete dissociation of telomerase from nucleoli at all stages of the cell cycle. Therefore, in normal cells, telomerase is sequestered from its potential substrate for most of the cell cycle, thus reducing its opportunity for access to chromatin. On the other hand, induced DNA double-stranded breaks cause the nucleolar association of telomerase in both primary and transformed cells and this sequestration was suggested to play a positive regulatory role in reducing the potential for inappropriate addition of telomeres to fragmented chromosomes.

In budding yeast, the phosphatase CDC14 is essential for inactivation of a mitotic kinase, i.e. cyclinB-cdk1 kinase (Garcia and Pillus, 1999; Visintin et al., 1999; Visintin and Amon, 2000; Shou et al., 2001; Geymonat et al., 2002). The inactivation of this kinase is a prerequisite for mitotic exit and cytokinesis in all organisms so far examined. During interphase, Cdc14 is inactive and localises to the nucleolus by binding to Net1, the same component that recruits the repressor protein Sir2 to rDNA (Section I.2.2). Following entry into mitosis, and specifically during anaphase when nucleoli segregate to daughter cells (Section I.2.4), Cdc14 is liberated from Net1, which remains in the nucleolus, and becomes active. The release of Cdc14 occurs presumably in response to a signal

transduction cascade initiated by the anaphase-promoting complex. In addition, the repressor protein Sir2 is also involved in monitoring appropriate progression through meiosis. Sir2 recruits a protein involved in the meiotic checkpoint to the nucleolus and thereby represses meiotic homologous recombination between rDNA (San-Segundo and Roeder, 1999).

I.2.5.3 Other functions of the Nucleolus

The convergence of silencing and nucleolar proteins supports the idea that the nucleolus and other silenced or heterochromatic regions of the nucleus may be privileged regulatory sites. This has further developed by considering the potential mechanism of cellular ageing. *Saccharomyces cerevisiae* divides asymmetrically and the finite number of daughters produced by a mother cell is the measure of life span (Guarente, 1997; Kennedy et al., 1997; Sinclair et al., 1997; Johnson et al., 1998). A mutant screen revealed a gain-of-function allele of another SIR repressor family member, the Sir4 gene. Sir4 forms a complex with Sir2 and Sir3 to form a transcription silencing complex at telomeres and the silent mating type loci (Section I.2.2). The effect of *SIR4-42* redirects the Sir complex to the nucleolus, which lengthens life span, and the same translocation of the Sir complex was observed to occur in aging wild-type cells to promote longevity.

Mutants in *SGS1*, the yeast homologue of the human gene which is defective in Werner Syndrome patients, cause a 60% shortening of life span. This yeast protein, like its human homologue, localises to the nucleolus and is involved in rDNA recombination (Marciniak et al., 1998; Carmo-Fonseca, 2002). Strikingly, examination of both sgs1 ageing mutant and aged wild type mother cells revealed that the normally single nucleolus becomes fragmented and enlarged and contains abundant extrachromosomal rDNA circles (ERC). These ERCs are believed to be formed by occasional intrachromosomal homologous

recombination between rDNA repeats and they accumulate to high levels in ageing cells (Sinclair and Guarente, 1997; Sinclair et al., 1997). Even though so far no ERCs have been observed in mammalian cells, ageing may be related to the regulation of proper recombination between rDNA within the nucleolus. This may partly explain the usage of only half of the rDNA copies, even at maximal transcriptional output. Because transcribed genes are generally more accessible, they are therefore probably subjected to considerably more damage than inactive ones and this may be particularly true for the most intensely transcribed rDNA genes. The silenced rDNA repeats may hence be a source of reserves to replace irretrievably damaged repeats. Alternatively, these silenced genes may be involved in other functions of the nucleolus.

Growing evidence suggests that viruses may target the nucleolus and its components to favour viral transcription, translation and perhaps alter the cell cycle in order to promote virus replication (Hiscox, 2002). Viruses can interact with the nucleolus by localising components to there that can then cause the endogenous nucleolar proteins to relocate to the cytoplasm during infection. It was recently suggested that this cytoplasmic redistribution of nucleolar proteins and subsequent virus-induced cell rupture may result in presenting antigens normally resident within the nucleolus to the immune system. This cascade of virus induced action is suggested to be the cause of autoimmune diseases such as scleroderma and systematic lupus erythematosus as a result of autoimmunity to nucleolin and fibrillarin, respectively.

Many nucleolar phenomena still remain to be explained. For example, what are the functions of those RNA species that localise within the nucleolus but have no obvious relationship with ribosomal biogenesis, namely, the SRP RNAs, tRNAs and RNase P RNAs? And what other RNA modifications can occur in the nucleolus such as

modification of the spliceosomal U6 snRNA? What are the reasons behind the relocalisation of extracellular growth factors into the nucleolus? What is the role played by those proteins specifically localised in the nucleolus of stem cells, but not of differentiated cells? What more does the nucleolus have in store for us to discover? To answer these questions in my PhD, I have followed two parallel approaches aimed at understanding more about the function of the nucleolus:

A. by investigating what protein components constitute the human nucleolus using a proteomic approach (Section I.3.1); and
B. by following how these proteins localise to the nucleolus and investigating the relationship between the assembly/disassembly of the subnucleolar domains as a function of activity, using high resolution light microscopy (Section I.3.2).

I.3 Analysis Tools

I.3.1 Proteomics

I.3.1.1 General

'Proteomics' is the large scale characterisation of the entire protein complement, i.e. the proteome, of a cell line, a tissue or even an organism (Roepstorff, 1997; Andersen and Mann, 2000; Pandey and Mann, 2000). Traditionally, it has been used to describe the analysis of protein components separated by 2-dimensional gel electrophoresis (Fey and Larsen, 2001). However, this method cannot easily determine the identities of proteins using traditional chemical method such as Edman protein sequencing. A method which is both sensitive and high-throughput for protein characterisation was in dire need during the 1990s and this method came after two important technical breakthroughs. The first one came from the field of mass spectrometry and the other from the field of molecular biology.

Mass spectrometry has traditionally been used to identify the mass of a small molecule with high accuracy. However it was not widely used for protein analysis until the emergence of two new ionization techniques that allow the production of gas phase molecular ions of protein fragments, or peptides, i.e. electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI; Section I.3.1.2). Currently it is possible to measure even a subpicomole amount of a peptide with a mass accuracy better than 0.01% (Smith, 2002). The current general scheme for protein characterisation involves the identification of peptides after protease digestion of a purified protein and the search for peptide matches from protein databases. At the same time, the development of the polymerase chain reaction made both sequencing the whole genome and large scale mRNA analysis possible. The protein information derived from the human genome and the mRNA data enables researchers to search not only the primary databases, such as SWISSPROT and, now more commonly, International Protein Index (IPI), but also the genome and expressed sequence tag (EST) databases by *in silico* translation (Roepstorff, 1997; Neubauer et al., 1998a; Fenyo, 2000; Pandey and Mann, 2000; Andersen et al., 2002).

These advances in protein analysis have resulted in a huge catalogue of protein components from biochemical complexes, organelles and even whole organism proteomes. Different approaches were developed to identify a set of proteins purified with a specific function in mind. For example, spliceosomes were purified using streptavidin affinity chromatography to isolate biotinylated pre-mRNA substrates after *in vitro* spliceosome assembly (Neubauer et al., 1998b; Rappsilber et al., 2002; Zhou et al., 2002b). Changes in protein composition during spliceosome activation were analysed by comparing the immunocomplexes from an *in vitro* splicing reaction and nuclear extract using antibodies raised against a specific splicing factor (Makarov et al., 2002). Similarly, antibodies raised against phosphotyrosine have been used to identify potential substrates in several receptor-

mediated signaling pathways by comparing the factors isolated from both unstimulated and growth factor-stimulated cell lysates (Conrads et al., 2002; Mann et al., 2002).

A high throughput approach to identify the interaction partners of a protein, commonly known as 'bait', is to express the protein fused with one or two cleavable high affinity tags (Rigaut et al., 1999). This approach is especially powerful in genetically amenable species, such as budding yeast. Large-scale interaction mapping has been performed to understand the ribosome maturation and DNA repair pathways (Gavin et al., 2002; Ho et al., 2002; Nissan et al., 2002). Once the composition of a complex is known, additional baits can be chosen from this known set of proteins and used to pull out other associated partners. By comparing the factors identified using different baits, this type of 'complex walking' approach reveals a series of snapshots of the change in composition of a multi-protein complex as it either functions or proceeds along a pathway (e.g. Section I.2.5.1) (Nissan et al., 2002; Shevchenko et al., 2002).

I.3.1.2 Ion source

By definition, a mass spectrometer consists of an ion source, a mass analyzer and a detector (Roepstorff, 1997; Andersen and Mann, 2000; Smith, 2002). Generated ions are deflected by a magnetic field according to their mass/charge ratios and are measured and registered by the mass analyzer and the detector, respectively. In 1988, coincidentally, two different ionization techniques were developed to generate gas phase ions from a liquid source, i.e. MALDI and ESI. Table I-2 summarises the differences between MALDI and ESI mass spectrometry.

In MALDI, a protein solution embedded in a co-precipitate of light absorbing matrix (usually α-cyano-4-hydroxycinnamic acid or dihydroxy benzoic acid) is irradiated in a

vacuum with a short pulse of UV light. Rapid sublimation of the analyte/matrix crystallites leads to the formation of protonated molecular ions in gas phase. In ESI, analyte solution is passed through a hypodermic needle head at high potential. Due to the high electric potential, the solution disperses into a mist of small, highly charged droplets (usually one proton per kDa) containing the analytes, which evaporate rapidly as the protonated molecular ions are released into gas phase. Due to the evaporation mechanism in this condition, there is only one analyte molecule per droplet formed at the needle tip. The introduction of the nanoelectrospray source allows as little as 0.5µl of analyte solution for a stable spray lasting up to 30 minutes, gives rise to improved sensitivity and longer analysis time (Mann and Wilm, 1995; Wilm and Mann, 1996).

	MALDI-MS	**ESI-MS**
Mass range (Da)	$400 - 2 \times 10^5$	$50 - 10^5$
Mass accuracy (peptide)	0.1-0.001%	0.05-0.008%
Sensitivity	Low fmol to pmol	Low fmol to pmol
Resolution (peptide)	500-10000	1000-2000
Tolerance to impurities	Good	Moderate
LC-MS capability	Offline	Online
MS/MS	Some	Yes
Cost	**$$**	**$$$$**
Mass analyzer (typical)	Time-of-flight (TOF)	Quadrupole mass filter

Table I-2 **Comparison of MALDI and ESI mass spectrometry.** LC-MS: liquid chromatography coupled mass spectrometry and MS/MS: tandem mass spectrometry commonly used for identification of amino acid sequence of peptides. Adapted from Mann and Wilm (1995).

I.3.1.3 Mass analyzer and spectrum detector

The advantage of mass spectrometry compared with all other methods in determining the mass of a molecule is its high accuracy and therefore the mass analyzer is arguably the crucial part of the technology. Proteomic measurements should mostly aim for high throughput, the ability to be coupled with other separation technologies such as liquid chromatography (LC), high sensitivity and large dynamic range and mass accuracy.

Currently 3 different types of mass analyzers are commonly used, namely Quadrupole Time-of-Flight (QTOF), Ion Trap (IT) or Fourier Transform Ion Cyclotron (FTIC). Table I-3 summarises the relative strengths of these mass analyzers. TOF analysis is basically a measure of the time required for a peptide ion to fly from the ion source to the detector, which varies according to the mass of the intact peptide. Based on the high mass accuracy of the analyzer, it is possible to deduce the composition of the peptide from reference to existing databases. The disadvantage of this technique is that it cannot identify proteins that have not previously been deposited in databases. Since this type of mass mapping (more commonly known as fingerprinting) requires a highly purified target protein, the technique is commonly used in conjunction with either one or two-dimensional gel electrophoresis to separate and purify proteins. Software is available to compare and analyse spots that vary between different gels (Fey and Larsen, 2001).

Instrument Type	Detection limit (moles)	Dynamic Range	Resolving Power	Mass Accuracy	Cost
IT	10^{-16}-10^{-14}	~10^2	~10^3	300-1000ppm	**$**
QTOF	10^{-15}-10^{-13}	~10^2	~10^4	5-40ppm	**$$**
FTIC	10^{-19}-10^{-15}	~10^3	~10^5	<1-10ppm	**$$$$**

Table I-3 **Mass spectrometry performance in proteome measurement.** IT: Ion trap; QTOF: Quadrupole Time-of-Flight; FTIC: Fourier Transform Ion Cyclotron. Adapted from Smith 2002.

Currently, it is more common to couple protein separation, through one or two liquid chromatography steps such as ion exchange chromatography and gradient reversed-phase, with an ESI tandem mass spectrometry analysis, i.e. LC-ESI-MS/MS. In LC-ESI-MS/MS, an electrospray is often coupled to ion traps and triple quadrupole instruments (Smith, 2002). This setup allows peptide sequencing by further generating fragment ion spectra, i.e. collision induced dissociation spectra (CID), of peptide precursor ions selected from the first mass spectrometer. The fragmentation of the precursor peptide results in a nested set of peaks differing by amino acid residue masses and the peptide primary sequence can

hence be reconstructed (Mann and Wilm, 1995). The advantage of speed, high sensitivity and the ability to work from unseparated peptide mixtures effectively makes traditional Edman sequencing obsolete. The general scheme for protein identification by this approach can be summarized as follows:

Protein recovery → *Enzymatic digestion* → *Liquid chromatography separation* → *Mass spectrometry* → *selected peak subjected to MS/MS* → *Database search from the mass/charge information* → *Peptide (and protein) identification*

However, the increasing complexity of the samples analysed, for example from a whole organism, requires strategies to increase proteome coverage. This includes computer control methods to minimize subsequent re-analysis of peptides, multiple analyses of the same sample with attention focused on different mass/charge ranges and the automated adjustment of the LC rate based on the number of peptides co-eluted. Recently, it was proposed to use both the information from separation (e.g. LC elution times) and the high mass measurement accuracy achieved in FTIC (which extends to at least 0.1 ppm) for peptide identification. This information from both accurate mass and time (AMT) forms a tag for each peptide (Conrads et al., 2000; Lipton et al., 2002; Smith et al., 2002). In the AMT tag strategy, peptides are first tentatively identified in multiple runs by using LC-MS/MS. The tentatively identified peptides are then validated by high accuracy measurement of the peptides' predicted mass at the same relative elution time in LC separation. This strategy has been used for the identification of the *Deinococcus radiodurans* proteome, covering ~61% of the open reading frames of which a quarter were previously assigned as either hypothetical, or conserved hypothetical (Lipton et al., 2002).

The large variation in relative protein abundance within a sample, which could be as high as 10 orders of magnitude in the case of human plasma, presents a major challenge for proteomics (Anderson and Anderson, 2002). The dynamic range can be improved by the

aforementioned approach including liquid chromatography and the AMT strategy. In addition, the methodology known as Dynamic Range Enhancement Applied to Mass Spectrometry (DREAMS) have been reported recently to solve this issue, e.g. by removing the most abundant ion peaks found in previously obtained spectra (Belov et al., 2001; Pasa-Tolic et al., 2002).

I.3.1.4 Quantitative proteomics

Because the main advantage of mass spectrometry is its mass accuracy, it is possible to add a quantitative dimension to proteomic experiments by adding reagents carrying an isotope with a molecular mass that is different from the isotope used in the control experiment (Flory et al., 2002). The chemically identical analytes with different stable isotope compositions can be differentiated by their mass differences and the signal intensities between the two forms can indicate their relative abundance ratio. Stable isotope tags have been introduced to samples post isolation via multiple techniques, including metabolic labeling using heavy salts, or amino acids (Conrads et al., 2002), enzymatically via transfer of ^{18}O from water (Yao et al., 2001), or more commonly via chemical reactions using isotope coded affinity tags (ICAT) that covalently label a particular amino acid such as cysteine (Gygi et al., 1999). Alternatively, stable isotope labeled amino acids such as ^{13}C arginine can be introduced in cells growing in a particular metabolic state, e.g. when transcription is inhibited before isolation (Ong et al., 2002).

I.3.2 Microscopy

Knowledge concerning the biochemical constituents in a structure is crucial to understand how and where they act within the cell. Moreover, the localisation and the quantitative behaviour of molecules in cells is a critical part of their functions. Therefore, it is important to understand how molecules behave, especially in live cells. As cells are three-dimensional

structures, it requires multiple images with a defined depth-of-field taken at regular focus intervals, i.e. optical sections, to represent the sample. However, only a small fraction of the light that is illuminating through the sample and entering the lens, originates from the focal plane where features are in sharp focus. Therefore, the object in focus is often obscured by a blurry background due to the 'out-of-focus' light.

I.3.2.1 Hardware microscopy equipment

There are two common ways used to solve the issue of the blurred image in biological samples, i.e. the laser scanning confocal microscope (LSM) and the wide-field microscope (WFM; Figure I-6). The principle of LSM is to use a focused laser to excite the sample point-by-point across the x-y plane and eliminate the out-of-focus light by imposing a

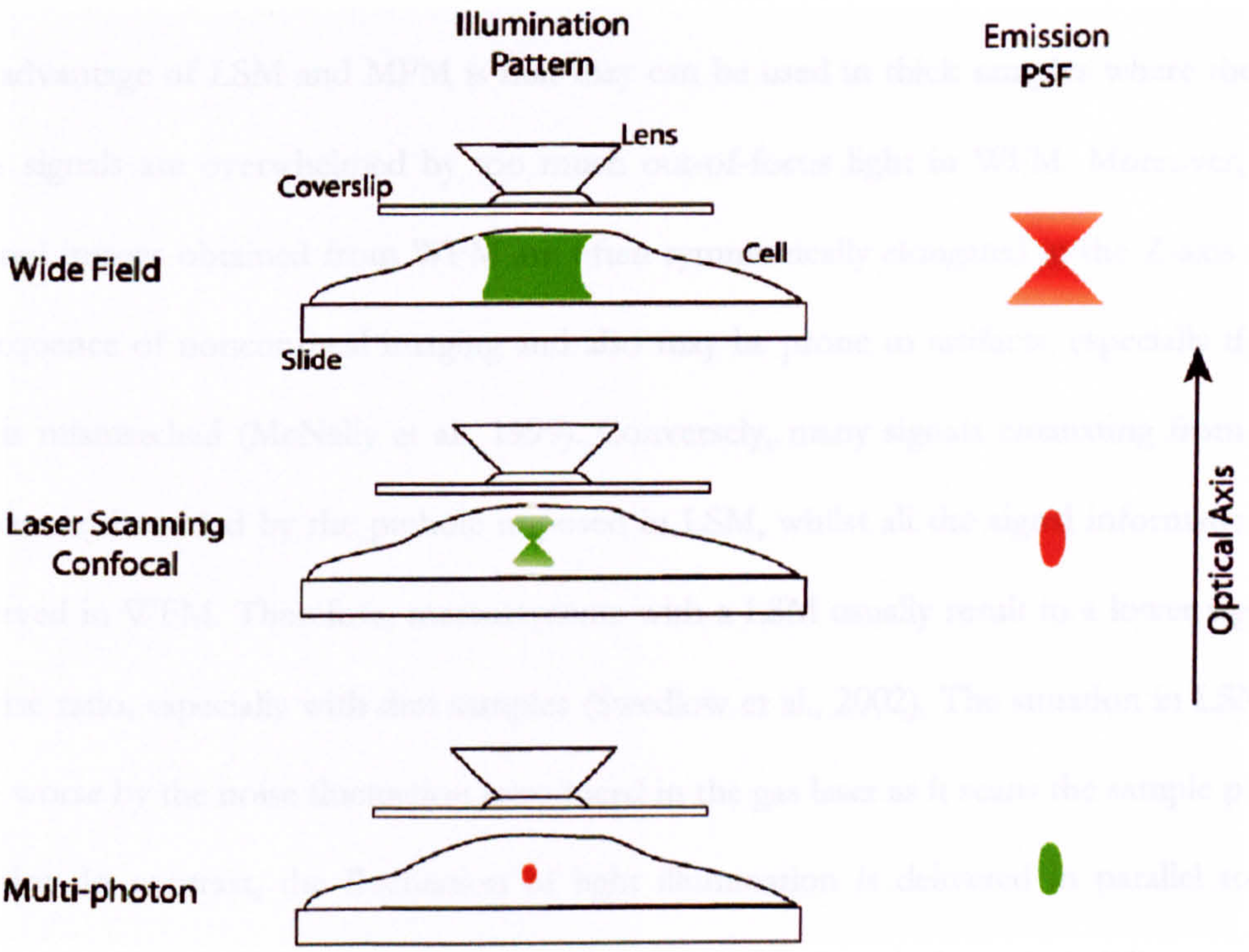


Figure I-6 **High resolution Light Microscopes.** The diagram shows a schematic diagram illustrating a typical layout of an objective lens, coverslip, cell and slide. The left hand side shows the volume of fluorescence excitation and the right hand side shows the corresponding fluorescence from each point source (Andrew et al., 2002).

pinhole before detecting the emitted fluorescence with a photomultiplier tube (Pawley, 1995; Andrews et al., 2002). Alternatively, a multiphoton microscope (MPM) can limit the laser-induced fluorescence excitation to a small volume by only allowing the absorption of two photons possessing half the energy, i.e. twice the usual excitation wavelength. As it requires high photon density at a focused point to excite the sample, out-of-focus fluorescence is negligible. In contrast, a WFM illuminates the whole sample simultaneously and records the emitted fluorescence with a charge-coupled device (CCD) camera (McNally et al., 1999; Andrews et al., 2002; Swedlow and Platani, 2002). To recover an estimate of the unblurred object from in-focus and out-of-focus signals, algorithms are developed to 'deconvolute' the image based on the knowledge of how a point source light behaves in the corresponding 3D image, i.e. the point spread function (PSF).

The advantage of LSM and MPM is that they can be used in thick samples where the in-focus signals are overwhelmed by too much out-of-focus light in WFM. Moreover, the restored images obtained from WFM are often symmetrically elongated in the Z-axis as a consequence of nonconfocal imaging and also may be prone to artifacts, especially if the PSF is mismatched (McNally et al., 1999). Conversely, many signals emanating from the sample are discarded by the pinhole imposed in LSM, whilst all the signal information is preserved in WFM. Therefore, measurements with a LSM usually result in a lower signal-to-noise ratio, especially with dim samples (Swedlow et al., 2002). The situation in LSM is made worse by the noise fluctuation introduced in the gas laser as it scans the sample point by point. In contrast, the fluctuation of light illumination is delivered in parallel to all regions of the sample simultaneously by WFM, eliminating pixel-to-pixel differences in delivered illumination.

Although the high intensity of illumination by a focused laser may not be advantageous to the cells in LSM, it may be useful in some quantitative experiments; e.g. photobleaching in a short period of time (Section I.3.2.3). In a live cell experiment, it may take a much longer exposure by the low illumination from a WFM light source in order to obtain the same signal intensity achieved by a shorter time interval, but higher powered laser pulse, in LSM (Andrews et al., 2002; Swedlow and Platani, 2002). Therefore, a compromise must be reached between the signal-to-noise ratio and the time required to excite the sample as both high intensity light and prolonged exposure are detrimental for cellular processes such as mitosis. In this thesis, a combination of both LSM and WFM has been employed, depending on the type of experiments as illustrated in Chapters IV and V.

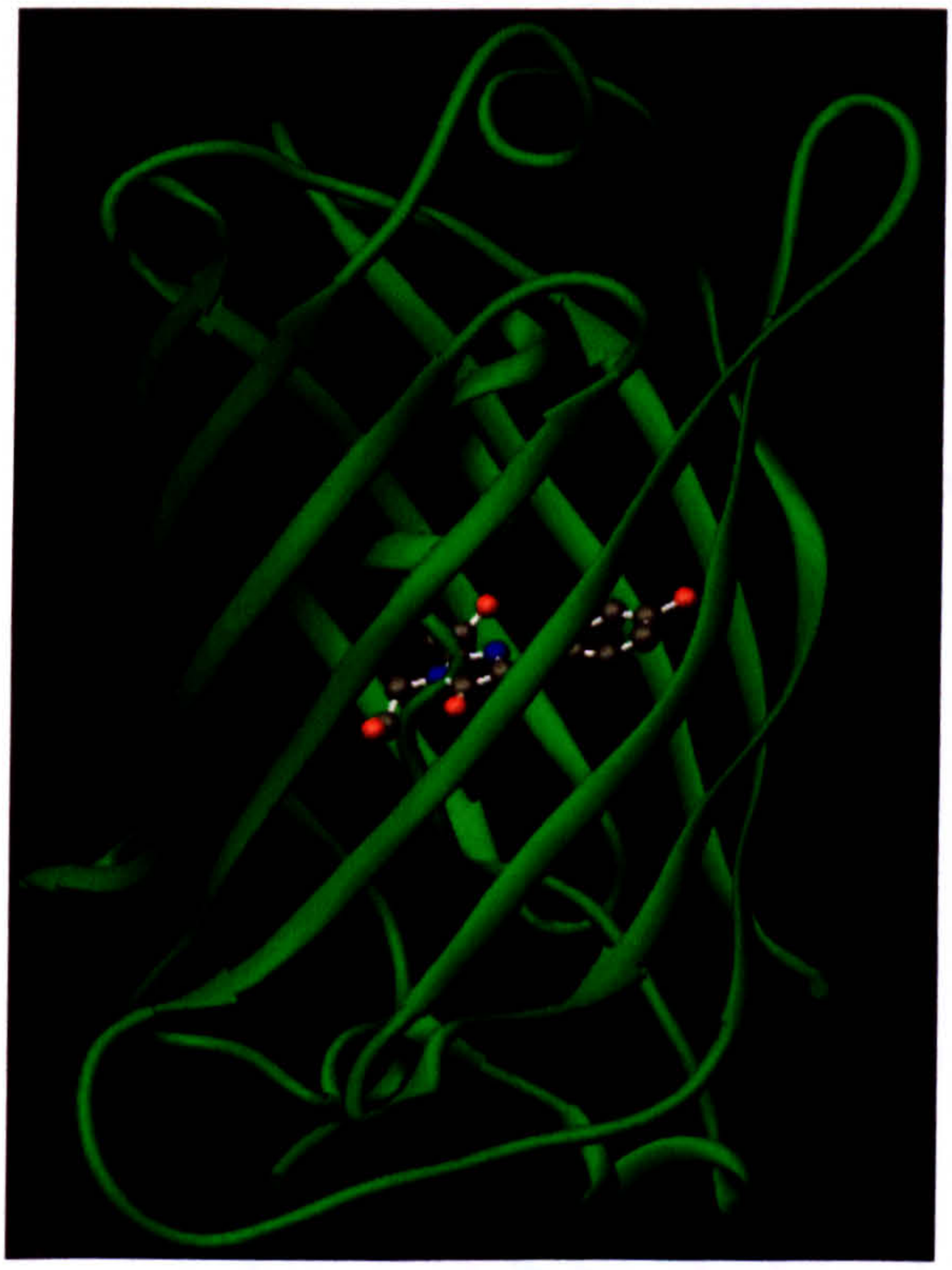


Figure I-7 **Crystal Structure of GFP and the sites of the fluorophore.** Ribbon diagram of GFP drawn from the wild-type crystal structure. The buried chromophore, which is responsible for GFP's luminescence, is shown in full atomic detail. (http://www.npaci.edu/online/v4.14/gfp.htm).

I.3.2.2 Live Cell Fluorophores

Traditionally, the localisation of a protein was investigated by immunofluorescence in which a protein is detected by the signal of a fluorophore that is coupled directly, or more often indirectly, via secondary antibodies, to a cognate antibody. The samples must therefore be fixed before binding the antibody. However, the cloning of the green fluorescent protein (GFP) from *Aequorea victoria* has changed the field of cell biology (Tsien, 1998). The chromophore is a *p*-hydroxybenzylideneimidazolinone

formed from residues 65-67 (Figure I-7), which are Ser-Tyr-Gly in the native protein, generated by cyclisation and oxidation steps after folding. GFP is an 11-stranded β-barrel threaded by an α helix that runs up the axis of the cylinder and bears the chromophore (Yang et al., 1996). The structural stability and relative resistance to protease cleavage make it an ideal tool to express as a fusion protein with the protein of interest for localisation studies, due to its ability to self-generate a chromophore (Misteli and Spector, 1997; Lippincott-Schwartz et al., 2001; van Roessel and Brand, 2002). Moreover, the localisation can now be followed in live cells, thus avoiding any potential fixation artifacts, although it must always be borne in mind that the GFP tag can potentially influence the properties of the protein it is fused to.

Apart from optimization of the codon usage and removal of cryptic splice sites within GFP, a series of site-directed mutants have been made to change the excitation and emission spectra of the chromophore (Figure I-8). These mutations enhance its brightness and improve folding at 37°C, based on the essential knowledge of the X-ray crystal structure (Tsien, 1998). Wild type GFP normally exists as a mixed population of chromophores, namely the neutral phenols and anionic phenolates, which produce a major

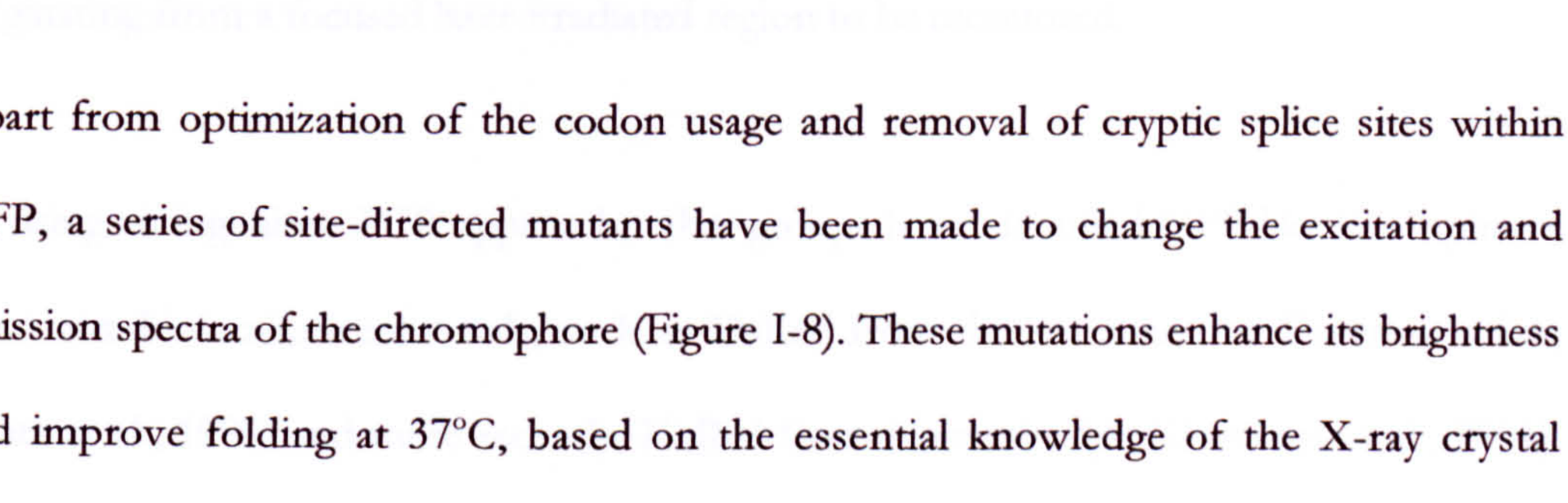


Figure I-8 **Emission and Excitation spectra of Fluorescent proteins. (A)** The emission and excitation spectra from ECFP, EGFP, EYFP, DsRed and HcRed are shown in blue, green, yellow, read and purple respectively. The figure was modified and compiled according to data obtained in Clontech.

397nm and a minor 475nm absorbance, respectively (Tsien, 1998). Upon intense illumination of the protein with 413 nm light, the chromophore population undergoes photoconversion and shifts the chromophore predominantly to the anionic form. This gives rise to an increase in the minor peak absorbance and consequently three fold increase in fluorescence emission upon excitation at 488nm. Patterson et al have recently generated a photoactivatable GFP whose minor absorbance peak was initially lower, such that the subsequent photoconversion by 413 nm irradiation produces an even greater increase in absorbance at the minor peak and ~100 times brighter emission when excited by 488 nm irradiation (Patterson and Lippincott-Schwartz, 2002). The selective photoactivation of a region of interest within cells using 413nm light allows the movement of protein originating from a focused laser-irradiated region to be monitored.

By using an ingenious PCR approach, other groups have identified GFP-like fluorophores from non-bioluminescent corals, such as DsRed from the sea anemone *Discosoma striata*, (Matz et al., 1999) and more recently, HcRed from *Heteractis crispa* (Gurskaya et al., 2001). Both DsRed and HcRed are red fluorescent proteins (RFPs; Figure I-8) and are potentially ideal for studying protein-protein interactions between GFP and RFP, similar to the way that has been used between cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), by fluorescence resonance energy transfer (FRET). This technique measures the degree of sensitised fluorescence emission from an acceptor, such as RFP, as a result of excitation of the donor, such as GFP (Truong and Ikura, 2001; van Roessel and Brand, 2002). One of the criteria for FRET is that the emission spectrum of the donor must overlap with the excitation spectrum of the acceptor. The energy transfer efficiency is inversely proportional to the 6^{th} power of the distance between the two pairs and therefore it is an ideal tool to study protein-protein interactions. So far, only the X-ray crystallographic structure of DsRed has been solved for the family of red fluorescent

proteins and the structure has a very similar 11 stranded β-barrel fold to the GFP with the chromophore formed between residues 66-68 (-Gln-Tyr-Gly-) (Wall et al., 2000). The red shift of the spectrum is caused by an additional oxidation step which forms an acylimine extension to the chromophore from neighbouring residues (Yarbrough et al., 2001). The original DsRed is very slow in folding and is a tetramer. Subsequent mutagenesis studies increased the rate of folding and monomerised the wild-type DsRed by reducing the monomer-monomer interactions (Campbell et al., 2002). Moreover, a mutant was generated such that there is a delay in the additional oxidation step responsible for the red shift (Terskikh et al., 2000). This mutant, known as 'Fluorescent timer', requires a relatively longer period of time for the red-shift to occur and hence provides a time window to change from green to red fluorescence, thereby providing a way to distinguish newly expressed protein from the old ones.

In addition, tagging fluorescent proteins (FPs) to different proteins/peptides increases their possible usage. For example, tagging FPs with the MS2-RNA binding protein or the lacI repressor enables researchers to study the movement of MS2-binding site-containing RNAs or a Lac operon cassettes integrated in DNA loci, respectively (Belmont and Straight, 1998; Bertrand et al., 1998). Systems have been established to study gene activity at the level of genome, mRNA expression and protein expression by a clever combination of these visualization tools (Tsukamoto et al., 2000). Addition of a degradation domain renders FP destabilised and can be used to directly correlate gene induction with a biochemical change (Li et al., 1998). A large-scale 'protein trap' strategy has tagged FP randomly, either to the 3' end, or within intron regions, of genes to identify their specific intracellular locations (Sutherland et al., 2001).

Another class of live cell fluorophores that can be genetically engineered to use in live cells is a reporter system called FlAsH (Griffin et al., 1998). This involves tagging of the protein of interest with a small motif (6-20 residues) that contains the sequence -Cys-Cys-X-X-Cys-Cys-, where the thiol groups of cysteine residues are arranged in a conformation to bind a membrane-permeant, nonfluorescent biarsenical derivative of fluorescien, i.e. 4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein. FlAsH becomes strongly green fluorescent when bound specifically to the tetracysteine motif. This system allows the target protein with a relatively short motif addition to be singled out inside live cells by staining with this small nonfluorescent dye. Recently, another biarsenical derivative of the red fluorescent resorufin, known as ReAsH, has been introduced (Gaietta et al., 2002). This has been used in parallel to label two temporally separated pools of connexins by administering the two reagents at different times. The ReAsH has the added advantage that it can photoconvert diaminobenzidine (DAB) to produce an electron dense product and hence allow the tagged protein to be imaged in both fluorescence and electron microscopy.

I.3.2.3 Quantitative Microscopy

Apart from FRET, a technique discussed in the last section that can provide information about the distance between two proteins, fluorescence microscopy can also measure protein dynamics quantitatively by using photobleaching techniques (Lippincott-Schwartz et al., 2001; Misteli, 2001; Phair and Misteli, 2001; Reits and Neefjes, 2001). Photobleaching is a photo-induced alteration of a fluorophore that extinguishes its fluorescence irreversibly. GFP is an ideal tool with which to study protein dynamics because the chromophore has a high fluorescence yield and it is resistant to photobleaching at low illumination. However, when excited with high illumination levels, the GFP fluorophore can be irreversibly photobleached and this was used to study the movement of non-bleached GFP fusion protein into the bleached areas (Figure I-9)

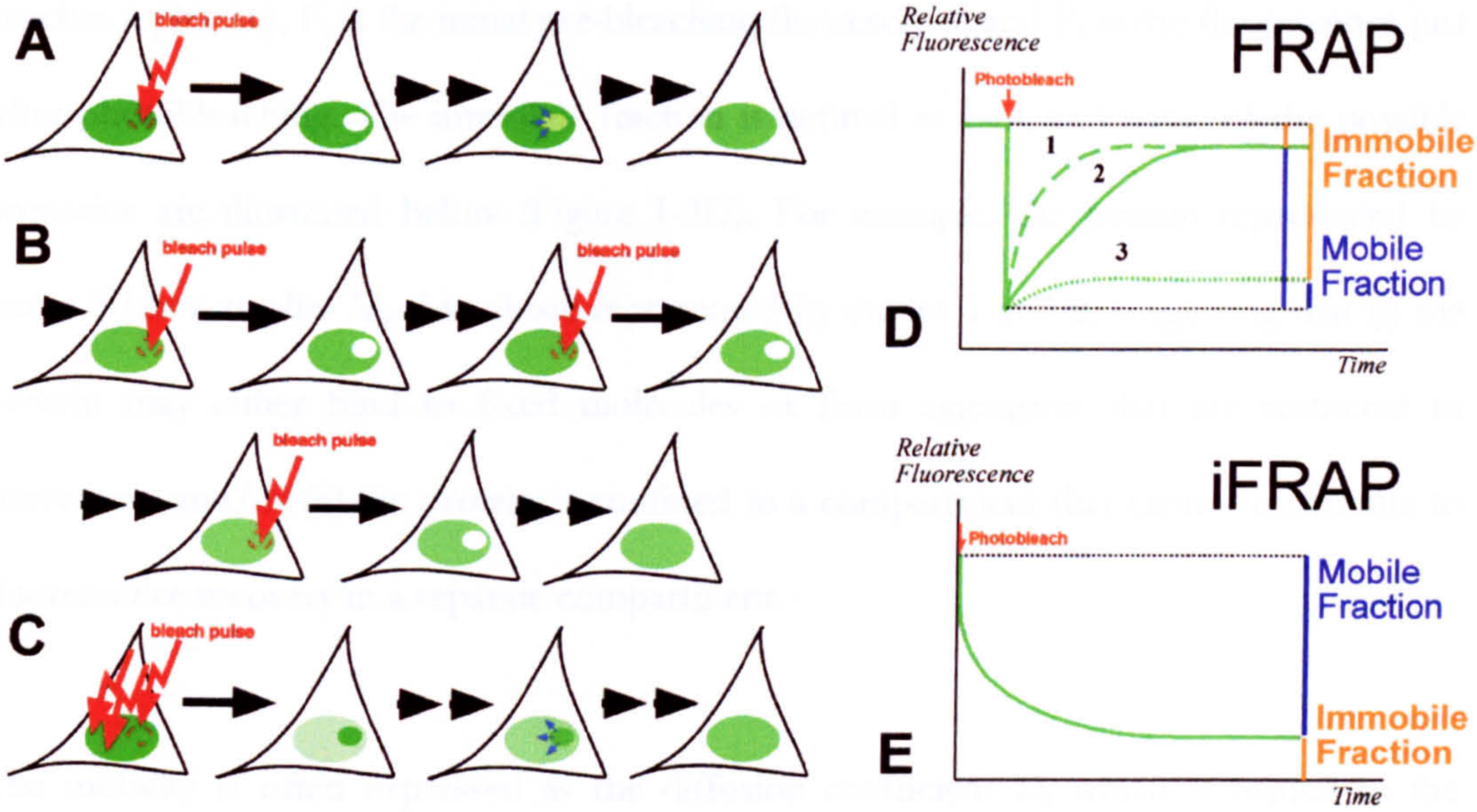


Figure I-9 **Microscopic experiments using photobleaching properties of GFP.** (A) FRAP, (B) FLIP and (C) iFRAP. (A) In FRAP, an area is photobleached once and the fluorescence recovery is monitored in that area while (B) FLIP is simply a repetitive form of FRAP but the fluorescence level outside the bleached area is monitored and (C) iFRAP is to monitor the loss of fluorescence of a small unbleached area when all other areas are bleached (see text for details). Both (D) FRAP and (E) iFRAP can provide a quantitative measurement and a qualitative description of the protein mobility, suggested by the shape of the curve. For example, proteins accounting for curve 1 have a higher diffusion coefficient compared with proteins accounting for curve 2 and 3 whereas proteins accounting for curve 3 may be more confined to a separate compartment resulting in their inability to exchange with other non-bleached fluorescent molecules to recover the fluorescence loss.

FLUORESCENCE RECOVERY AFTER PHOTOBLEACHING (FRAP)

After photobleaching of the fluorescent molecule in a small area of the cell, FRAP monitors how fast/slow the fluorescence within the designated area recovers based on the diffusion of non-bleached fluorescent molecules into the bleached area (Figure I-9A). This method was initially used in the 1970s using fluorophores coupled to proteins and lipids to study their movement within the plasma membrane and is currently used extensively to study protein dynamics using GFP-fusion proteins. Two parameters can be deduced from FRAP: (a) the mobile/immobile fraction and (b) the rate of diffusion time τ_D. The mobile fraction (M_f) is defined as follows:

$$M_f = \frac{F_\infty - F_0}{F_i - F_0} \qquad \textit{(Equation 1.1)}$$

where F_∞ is the fluorescence in the bleached region after full recovery (when the curve reaches a plateau), F_i is the initial pre-bleaching fluorescence and F_0 is the fluorescence just after photobleaching. The immobile fraction is defined as 1-M_f and some of the possible scenarios are illustrated below (Figure I-9D). For example, the protein represented by curve 3 has a smaller M_f than those represented by curves 1 and 2, suggesting that (i) the protein may either bind to fixed molecules or form aggregates that are restricted in movement and/or (ii) the protein is confined to a compartment that cannot contribute to fluorescence recovery in a separate compartment.

The mobility is often expressed as the diffusion coefficient *D*, which is related to the diffusion time τ_D and gives a measure of the rate of protein movement in the absence of either flow or active transport:

$$D = \frac{kT}{6\pi\eta R} \qquad \textit{(Equation 1.2)}$$

where T is the absolute temperature, η is the viscosity of the solution, k is the Boltzmann constant and R is the hydrodynamic radius of the particle. Therefore, when the viscosity and temperature are unchanged, *D* is mainly determined by the radius of the molecule and is roughly the inverse of the cube root of molecular mass, assuming that the object has a spherical shape. *D (in the unit of $\mu m^2 s^{-1}$)* reflects the mean square displacement that a protein explores through a random walk over time. For example, the increase in *D* (curve 1 compared with curve 2 in Figure I-9D) may suggest either a loss in environment viscosity or a transient loss of interaction with either large or fixed molecules. To calculate the value of *D*, the fluorescence recovery data are usually fit to a variant of equation 1.2 by a non-linear least square technique using mathematical software such as SAAM II or Berkeley Madonna.

FLUORESCENCE LOSS IN PHOTOBLEACHING (FLIP)

Although FLIP is not quantitative, it provides a simple way to assess the continuity of compartments. FLIP is similar to FRAP; but instead of photobleaching once, the photobleached area is repetitively bleached and the fluorescence level outside the area is monitored (Figure I-9B). If the molecules are freely diffusible between two compartments, the continual movement of molecules exchanging between the unbleached and the bleached compartments will deplete the pool of fluorescent molecules in the unbleached compartment and therefore result in a decrease in fluorescence within the unbleached compartment.

INVERSE FRAP(IFRAP)

Recently, a modified bleaching method known as iFRAP was introduced by the Misteli group to study the elongation rate of RNA polymerase I (Dundr et al, 2002). In an iFRAP experiment, the entire cell, apart from the area of interest is photobleached such that a snapshot of fluorescent molecules at the time of bleaching is obtained and the subsequent loss of fluorescence over time in the non-bleached area allows the corresponding dissociation of the fluorescent molecules in that area represented by a constant, k_{off} (Figure I-9C). Figure I-9E shows a typical fluorescence decay curve from an iFRAP experiment and indicates the corresponding mobile and immobile fractions. To calculate the dissociation constant, the decay curve is fitted to a defined model based on our knowledge of the system analysed, such as the stochiometry of the complex.

I.3.2.4 Other methods and microscopies

Apart from these photobleaching methods and FRET, fluorescence correlation spectroscopy (FCS) is becoming more popular as it is now possible to define a very small sample volume (~1 femotolitre) using confocal microscopy. FCS provides an alternative

way to measure diffusion coefficients by monitoring the fluctuations in photons resulting from the movement of fluorescent molecules between the defined volume and the surroundings (Van Craenenbroeck and Engelborghs, 2000). Other than GFP and the FlAsH system, recent successes include the production of single chain antibodies that can be fused with an organelle-targeting sequence to label a compartment of interest (Farinas and Verkman, 1999). Continual development of software and hardware, such as the 4-Pi microscope and the ultrafast dynamics microscope increase spatial resolution by factors of ~2-5 fold over the theoretical limit and certainly hold a promise to reveal greater details of the movement and localisation of proteins within subnuclear compartments (Nagorni and Hell, 1998; Klar et al., 2000; Thomann et al., 2002).

I.4 Aim of the Thesis

The aim of my thesis is to present the results I collected during my PhD studies on the following three aspects of a subnuclear compartment – i.e. the Nucleolus – with a hope that the principles learnt from this particular organelle can be extended to our understanding of other subnuclear compartments (Section I.1.2-I.1.3):

I.4.1 Characterisation of Nucleolar Proteome

As mentioned in Section I.2.5, so far we have learnt very little about the machineries involved in ribosome assembly in human cells, even though the nucleolus was identified as a 'ribosome factory' 40 years ago. Even less is known about what other roles may be played by the nucleolus. Our increasing knowledge of the functions in the nucleolus has mostly been initiated by fortuitous discoveries. The apparent question is, "what makes up the human nucleolus?". To answer this question, our group initiated 5 years ago a proteomic approach to systematically analyse the protein components of this organelle. When I joined the group in October 1999, the protocol for the isolation of nucleoli had been optimized

and an initial round of mass spectrometry analysis had confirmed its purity. However, even at that time, there were a lot of novel/uncharacterised proteins identified whose localisations remain unknown and hence my first goal was to test the localisation of these putative nucleolar factors by tagging them with GFP (Chapter III). The study of one of these novel factors subsequently led to the identification of a novel nucleolar targeting pathway (Chapter V). At the same time, I compiled a catalogue of all known nucleolar proteins identified since 1978, when the age of molecular cloning began, to help classifying the factors identified in the project (Chapter III).

As the project evolved, it became apparent that there exist at least 300-400 different proteins in the human nucleolus and a database was set up to archive both the primary information obtained from mass spectrometry and the secondary information derived from subsequent bioinformatics searches (Chapter III and http://www.dundee.ac.uk/lifesciences/lamonddatabase). Out of the ~400 proteins, ~30% of the proteins identified are classified as novel or uncharacterised. But because of the wealth of the secondary information, such as mRNA tissue distribution and homologues in other species, I attempted to use this derived data to classify the function of these novel factors and their potential relation to ribosome biogenesis (Chapter III).

I.4.2 Understanding of Subnucleolar domain dynamics

Although different subnucleolar domains have been studied before, it has so far not been studied in live human cells how these subnucleolar domains move in relation to one another, especially during their mitotic breakdown and reassembly. With this in mind, I have established 24 separate HeLa cell lines stably expressing one or more FP-tagged markers corresponding to the 3 morphologically distinct domains within the nucleolus, i.e. the FC, the DFC and the GC (Section I.2.4) and imaged nuclei from these cells from

prophase to early G1 (Chapter IV). These studies revealed that the breakdown and the assembly of the nucleolus may be initiated by the respective dissociation and recruitment of RNA polymerase I to the NORs. This is surprising, because it was long thought that RNA polymerase I remains associated with the chromosomes during different stages of mitosis but, in Chapter IV, I have shown that there exists a period between prophase and metaphase when RNA polymerase I dissociates from the chromosomes. Moreover, I have established an *in vivo* technique to label one or more nucleoli non-invasively to study their dynamics in different metabolic states and, together with other results, I propose a plausible mechanism for how the number and size of the nucleoli are determined. This provides a testing ground to understand why some malignant cells have extremely large nucleoli.

I.4.3 More than one route to Nucleoli

Another obvious question in studying nucleolar dynamics is, "how do proteins know where the nucleolus is and which part of the nucleolus they should go to?" In chapter V, I have demonstrated that, even though two proteins, namely fibrillarin and NHPX, localise in the same subnucleolar domain, they move to the nucleolus using apparently different mechanisms after their nuclear entry. NHPX transiently travels to the splicing speckles before accumulating in the nucleolus of both primary and transformed cell lines. This pathway is RNA polymerase II transcription-dependent and may link to multiple RNA pathways within the cell.

I.5 References

Aitchison, J.D., and M.P. Rout. 2000. The road to ribosomes. Filling potholes in the export pathway. *J Cell Biol.* 151:F23-6.

Akhmanova, A., T. Verkerk, A. Langeveld, F. Grosveld, and N. Galjart. 2000. Characterisation of transcriptionally active and inactive chromatin domains in neurons. *J Cell Sci.* 113 Pt 24:4463-74.

Alexander, S.P., and C.L. Rieder. 1991. Chromosome motion during attachment to the vertebrate spindle: initial saltatory-like behavior of chromosomes and quantitative analysis of force production by nascent kinetochore fibers. *J Cell Biol.* 113:805-15.

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. *Curr Biol.* 12:1-11.

Andersen, J.S., and M. Mann. 2000. Functional genomics by mass spectrometry. *FEBS Lett.* 480:25-31.

Anderson, N.L., and N.G. Anderson. 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 1:845-67.

Andrews, P.D., I.S. Harper, and J.R. Swedlow. 2002. To 5D and beyond: quantitative fluorescence microscopy in the postgenomic era. *Traffic.* 3:29-36.

Belmont, A.S., and A.F. Straight. 1998. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends Cell Biol.* 8:121-4.

Belov, M.E., G.A. Anderson, N.H. Angell, Y. Shen, N. Tolic, H.R. Udseth, and R.D. Smith. 2001. Dynamic range expansion applied to mass spectrometry based on data-dependent selective ion ejection in capillary liquid chromatography fourier transform ion cyclotron resonance for enhanced proteome characterization. *Anal Chem.* 73:5052-60.

Bertrand, E., P. Chartrand, M. Schaefer, S.M. Shenoy, R.H. Singer, and R.M. Long. 1998. Localization of ASH1 mRNA particles in living yeast. *Mol Cell.* 2:437-45.

Birnstiel, M., and M. Chipchase. 1970. The nucleolus: pacemaker of the cell. *Sci J.* 6:41-48.

Birnstiel, M., and B. Hyde. 1963. Protein synthesis by isolated pea nucleoli. *J Cell Biol.* 18:41-50.

Boudonck, K., L. Dolan, and P.J. Shaw. 1999. The movement of coiled bodies visualized in living plant cells by the green fluorescent protein. *Mol Biol Cell.* 10:2297-307.

Bowman, L.H., W.E. Goldman, G.I. Goldberg, M.B. Hebert, and D. Schlessinger. 1983. Location of the initial cleavage sites in mouse pre-rRNA. *Mol Cell Biol.* 3:1501-10.

Boyle, S., S. Gilchrist, J.M. Bridger, N.L. Mahy, J.A. Ellis, and W.A. Bickmore. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet.* 10:211-9.

Brown, D.D., and J.B. Gurdon. 1964. Absence of ribosomal RNA synthesis in the anucleolate mutant of Xenopus Laevis. *Proc Natl Acad Sci U S A.* 51:139-46.

Brown, J.D. 2001. Ribosome biogenesis: stripping for AAAction? *Curr Biol.* 11:R710-2.

Brown, R. 1833. Observations on the organs and mode of fecundation in Orchideae and Asclepiadeae. *Transactions of the Linnean Society of London.* 16:685-743.

Burke, B., and J. Ellenberg. 2002. Remodelling the walls of the nucleus. *Nat Rev Mol Cell Biol.* 3:487-97.

Cajal, S. 1903. Un sencillo metodo de coloracion seletiva del reticulo protoplasmatico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. *Trab. Lab. Invest. Biol. (Madrid).* 2:129-221.

Campbell, R.E., O. Tour, A.E. Palmer, P.A. Steinbach, G.S. Baird, D.A. Zacharias, and R.Y. Tsien. 2002. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A.* 99:7877-82.

Carmo-Fonseca, M. 2002. The contribution of nuclear compartmentalization to gene regulation. *Cell.* 108:513-21.

Carmo-Fonseca, M., C. Cunha, N. Custodio, C. Carvalho, P. Jordan, J. Ferreira, and L. Parreira. 1996. The topography of chromosomes and genes in the nucleus. *Exp Cell Res.* 229:247-52.

Carmo-Fonseca, M., L. Mendes-Soares, and I. Campos. 2000. To be or not to be in the nucleolus. *Nat Cell Biol.* 2:E107-12.

Carmo-Fonseca, M., M. Platani, and J.R. Swedlow. 2002. Macromolecular mobility inside the cell nucleus. *Trends Cell Biol.* 12:491-5.

Carvalho, T., F. Almeida, A. Calapez, M. Lafarga, M.T. Berciano, and M. Carmo-Fonseca. 1999. The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. *J Cell Biol.* 147:715-28.

Chen, D., and S. Huang. 2001. Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. *J Cell Biol.* 153:169-76.

Cheutin, T., M.F. O'Donohue, A. Beorchia, M. Vandelaer, H. Kaplan, B. Defever, D. Ploton, and M. Thiry. 2002. Three-dimensional organization of active rRNA genes within the nucleolus. *J Cell Sci.* 115:3297-307.

Conrads, T.P., G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, and R.D. Smith. 2000. Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem.* 72:3349-54.

Conrads, T.P., H.J. Issaq, and T.D. Veenstra. 2002. New tools for quantitative phosphoproteome analysis. *Biochem Biophys Res Commun.* 290:885-90.

Cremer, T., and C. Cremer. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet.* 2:292-301.

Darzacq, X., B.E. Jady, C. Verheggen, A.M. Kiss, E. Bertrand, and T. Kiss. 2002. Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *Embo J.* 21:2746-56.

Derenzini, M., D. Trere, A. Pession, L. Montanaro, V. Sirri, and R.L. Ochs. 1998. Nucleolar function and size in cancer cells. *Am J Pathol.* 152:1291-7.

Dietzel, S., and A.S. Belmont. 2001. Reproducible but dynamic positioning of DNA in chromosomes during mitosis. *Nat Cell Biol.* 3:767-70.

Dragon, F., J.E. Gallagher, P.A. Compagnone-Post, B.M. Mitchell, K.A. Porwancher, K.A. Wehner, S. Wormsley, R.E. Settlage, J. Shabanowitz, Y. Osheim, A.L. Beyer, D.F. Hunt, and S.J. Baserga. 2002. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature.* 417:967-70.

Dundr, M., U. Hoffmann-Rohrer, Q. Hu, I. Grummt, L.I. Rothblum, R.D. Phair, and T. Misteli. 2002. A kinetic framework for a mammalian RNA polymerase in vivo. *Science.* 298:1623-6.

Dundr, M., and T. Misteli. 2001. Functional architecture in the cell nucleus. *Biochem J.* 356:297-310.

Dundr, M., T. Misteli, and M.O. Olson. 2000. The dynamics of postmitotic reassembly of the nucleolus. *J Cell Biol.* 150:433-46.

Eichler, D.C., and N. Craig. 1994. Processing of eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol.* 49:197-239.

Farinas, J., and A.S. Verkman. 1999. Receptor-mediated targeting of fluorescent probes in living cells. *J Biol Chem.* 274:7603-6.

Fatica, A., and D. Tollervey. 2002. Making ribosomes. *Curr Opin Cell Biol.* 14:313-8.

Fenyo, D. 2000. Identifying the proteome: software tools. *Curr Opin Biotechnol.* 11:391-5.

Fey, S.J., and P.M. Larsen. 2001. 2D or not 2D. Two-dimensional gel electrophoresis. *Curr Opin Chem Biol.* 5:26-33.

Flory, M., T. Griffin, D. Martin, and R. Aebersold. 2002. Advances in quantitative proteomics using stable isotope tag. *Trends in Biotechnology.* 20:S23-S29.

Fox, A.H., Y.W. Lam, A.K. Leung, C.E. Lyon, J. Andersen, M. Mann, and A.I. Lamond. 2002. Paraspeckles. A novel nuclear domain. *Curr Biol.* 12:13-25.

Franke, W.W. 1988. Matthias Jacob Schleiden and the definition of the cell nucleus. *Eur J Cell Biol.* 47:145-56.

Gadal, O., D. Strauss, J. Braspenning, D. Hoepfner, E. Petfalski, P. Philippsen, D. Tollervey, and E. Hurt. 2001a. A nuclear AAA-type ATPase (Rix7p) is required for biogenesis and nuclear export of 60S ribosomal subunits. *Embo J.* 20:3695-704.

Gadal, O., D. Strauss, J. Kessl, B. Trumpower, D. Tollervey, and E. Hurt. 2001b. Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol Cell Biol.* 21:3405-15.

Gaietta, G., T.J. Deerinck, S.R. Adams, J. Bouwer, O. Tour, D.W. Laird, G.E. Sosinsky, R.Y. Tsien, and M.H. Ellisman. 2002. Multicolor and electron microscopic imaging of connexin trafficking. *Science.* 296:503-7.

Gall, J.G. 2000. Cajal bodies: the first 100 years. *Annu Rev Cell Dev Biol.* 16:273-300.

Gall, J.G., M. Bellini, Z. Wu, and C. Murphy. 1999. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol Biol Cell.* 10:4385-402.

Garcia, S.N., and L. Pillus. 1999. Net results of nucleolar dynamics. *Cell.* 97:825-8.

Gasser, S.M. 2001. Positions of potential: nuclear organization and gene expression. *Cell.* 104:639-42.

Gautier, T., M. Robert-Nicoud, M.N. Guilly, and D. Hernandez-Verdun. 1992. Relocation of nucleolar proteins around chromosomes at mitosis. A study by confocal laser scanning microscopy. *J Cell Sci.* 102 (Pt 4):729-37.

Gavin, A.C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature.* 415:141-7.

Gebrane-Younes, J., N. Fomproix, and D. Hernandez-Verdun. 1997. When rDNA transcription is arrested during mitosis, UBF is still associated with non-condensed rDNA. *J Cell Sci.* 110 (Pt 19):2429-40.

Geymonat, M., S. Jensen, and L.H. Johnston. 2002. Mitotic exit: the Cdc14 double cross. *Curr Biol.* 12:R482-84.

Gilbert, N., L. Lucas, C. Klein, M. Menager, N. Bonnet, and D. Ploton. 1995. Three-dimensional co-location of RNA polymerase I and DNA during interphase and mitosis by confocal microscopy. *J Cell Sci.* 108 (Pt 1):115-25.

Godfrey, K. 1983. Compartmental Models and Their Application. Academic Press, New York.

Gorlich, D., and U. Kutay. 1999. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol.* 15:607-60.

Granboulan, N., and P. Granboulan. 1965. Cytochemie ultrastructurale du nucleole: II etudedes sites de synthese du RNA dans le nucleole et le noyau. *Exp Cell Res.* 38:604-19.

Griffin, B.A., S.R. Adams, and R.Y. Tsien. 1998. Specific covalent labeling of recombinant protein molecules inside live cells. *Science.* 281:269-72.

Guarente, L. 1997. Link between aging and the nucleolus. *Genes Dev.* 11:2449-55.

Gurskaya, N.G., A.F. Fradkov, A. Terskikh, M.V. Matz, Y.A. Labas, V.I. Martynov, Y.G. Yanushevich, K.A. Lukyanov, and S.A. Lukyanov. 2001. GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett.* 507:16-20.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* 17:994-9.

Haaf, T., D.L. Hayman, and M. Schmid. 1991. Quantitative determination of rDNA transcription units in vertebrate cells. *Exp Cell Res.* 193:78-86.

Hadjiargyrou, M., K.A. McDowell, and A.S. Henderson. 1994. A transfected human ribosomal RNA gene is present in the nucleolus of human cells. *Cytogenet Cell Genet.* 66:58-62.

Hatanaka, M. 1990. Discovery of the nucleolar targeting signal. *Bioessays.* 12:143-8.

Heitz, E. 1931. Nukleolen und chromosomen in der gattung. *Vicia Planta.* 15:495-505.

Hernandez-Verdun, D., P. Roussel, and J. Gebrane-Younes. 2002. Emerging concepts of nucleolar assembly. *J Cell Sci.* 115:2265-70.

Hiscox, J.A. 2002. The nucleolus--a gateway to viral infection? *Arch Virol.* 147:1077-89.

Ho, J.H., G. Kallstrom, and A.W. Johnson. 2000. Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *J Cell Biol.* 151:1057-66.

Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, and M. Tyers. 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature.* 415:180-3.

Holaska, J.M., K.L. Wilson, and M. Mansharamani. 2002. The nuclear envelope, lamins and nuclear assembly. *Curr Opin Cell Biol.* 14:357-64.

Hozak, P., P.R. Cook, C. Schofer, W. Mosgoller, and F. Wachtler. 1994. Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells. *J Cell Sci.* 107 (Pt 2):639-48.

Huang, S. 2000. Review: perinucleolar structures. *J Struct Biol.* 129:233-40.

Huang, S. 2002. Building an efficient factory: where is pre-rRNA synthesized in the nucleolus? *J Cell Biol.* 157:739-41.

Huang, S., T.J. Deerinck, M.H. Ellisman, and D.L. Spector. 1997. The dynamic organization of the perinucleolar compartment in the cell nucleus. *J Cell Biol.* 137:965-74.

Hurt, E., S. Hannus, B. Schmelzl, D. Lau, D. Tollervey, and G. Simos. 1999. A novel in vivo assay reveals inhibition of ribosomal nuclear export in ran-cycle and nucleoporin mutants. *J Cell Biol.* 144:389-401.

Iborra, F.J., and P.R. Cook. 2002. The interdependence of nuclear structure and function. *Curr Opin Cell Biol.* 14:780-5.

Intine, R.V., M. Dundr, T. Misteli, and R.J. Maraia. 2002. Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs. *Mol Cell.* 9:1113-23.

Jarrous, N., R. Reiner, D. Wesolowski, H. Mann, C. Guerrier-Takada, and S. Altman. 2001. Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P. *Rna.* 7:1153-64.

Jarrous, N., J.S. Wolenski, D. Wesolowski, C. Lee, and S. Altman. 1999. Localization in the nucleolus and coiled bodies of protein subunits of the ribonucleoprotein ribonuclease P. *J Cell Biol.* 146:559-72.

Johnson, F.B., R.A. Marciniak, and L. Guarente. 1998. Telomeres, the nucleolus and aging. *Curr Opin Cell Biol.* 10:332-8.

Jordan, E.G., and J.H. McGovern. 1981. The quantitative relationship of the fibrillar centres and other nucleolar components to changes in growth conditions, serum deprivation and low doses of actinomycin D in cultured diploid human fibroblasts (strain MRC-5). *J Cell Sci.* 52:373-89.

Jordan, P., M. Mannervik, L. Tora, and M. Carmo-Fonseca. 1996. In vivo evidence that TATA-binding protein/SL1 colocalizes with UBF and RNA polymerase I when rRNA synthesis is either active or inactive. *J Cell Biol.* 133:225-34.

Karpen, G.H., J.E. Schaefer, and C.D. Laird. 1988. A Drosophila rRNA gene located in euchromatin is active in transcription and nucleolus formation. *Genes Dev.* 2:1745-63.

Kennedy, B.K., M. Gotta, D.A. Sinclair, K. Mills, D.S. McNabb, M. Murthy, S.M. Pak, T. Laroche, S.M. Gasser, and L. Guarente. 1997. Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in S. cerevisiae. *Cell.* 89:381-91.

Kiss, A.M., B.E. Jady, X. Darzacq, C. Verheggen, E. Bertrand, and T. Kiss. 2002. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res.* 30:4643-9.

Klar, T.A., S. Jakobs, M. Dyba, A. Egner, and S.W. Hell. 2000. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc Natl Acad Sci U S A.* 97:8206-10.

Klein, J., and I. Grummt. 1999. Cell cycle-dependent regulation of RNA polymerase I transcription: the nucleolar transcription factor UBF is inactive in mitosis and early G1. *Proc Natl Acad Sci U S A.* 96:6096-101.

Kobema, K., J. Malinsky, A. Pliss, M. Masata, J. Vecerova, M. Fialova, J. Bednar, and I. Raska. 2002. Ribosomal genes in focus: new transcripts label the dense fibrillar components and form clusters indicative of "Christmas trees" in situ. *J Cell Biol.* 157:743-8.

Kumar, A., and J.R. Warner. 1972. Characterization of ribosomal precursor particles from HeLa cell nucleoli. *J Mol Biol.* 63:233-246.

Labella, T., and D. Schlessinger. 1989. Complete human rDNA repeat units isolated in yeast artificial chromosomes. *Genomics.* 5:752-60.

Lafontaine, D.L., and D. Tollervey. 1998. Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci.* 23:383-8.

Lafontaine, J. 1968. Structural commponents of the nucleus in mitotic plant cells. *In* The Nucleus. A. Dalton and F. Hagenau, editors. Academic Press, New York. 152-96.

Lamond, A.I., and W.C. Earnshaw. 1998. Structure and function in the nucleus. *Science.* 280:547-53.

Lang, I., M. Scholz, and R. Peters. 1986. Molecular mobility and nucleocytoplasmic flux in hepatoma cells. *J Cell Biol.* 102:1183-90.

Lange, T.S., and S.A. Gerbi. 2000. Transient nucleolar localization Of U6 small nuclear RNA in Xenopus Laevis oocytes. *Mol Biol Cell.* 11:2419-28.

Lewis, J.D., and D. Tollervey. 2000. Like attracts like: getting RNA processing together in the nucleus. *Science.* 288:1385-9.

Li, X., X. Zhao, Y. Fang, X. Jiang, T. Duong, C. Fan, C.C. Huang, and S.R. Kain. 1998. Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem.* 273:34970-5.

Lippincott-Schwartz, J., E. Snapp, and A. Kenworthy. 2001. Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol.* 2:444-56.

Lipton, M.S., L. Pasa-Tolic, G.A. Anderson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y. Shen, E. Stritmatter, N. Tolic, H.R. Udseth, A. Venkateswaran, K.K. Wong, R. Zhao, and R.D. Smith. 2002. Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc Natl Acad Sci U S A.* 99:11049-54.

Lohrum, M.A., M. Ashcroft, M.H. Kubbutat, and K.H. Vousden. 2000. Identification of a cryptic nucleolar-localization signal in MDM2. *Nat Cell Biol.* 2:179-81.

Lyon, C.E., K. Bohmann, J. Sleeman, and A.I. Lamond. 1997. Inhibition of protein dephosphorylation results in the accumulation of splicing snRNPs and coiled bodies within the nucleolus. *Exp Cell Res.* 230:84-93.

Lyon, C.E., and A.I. Lamond. 2000. The nucleolus. *Curr Biol.* 10:R323.

Mahy, N.L., P.E. Perry, and W.A. Bickmore. 2002a. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol.* 159:753-63.

Mahy, N.L., P.E. Perry, S. Gilchrist, R.A. Baldock, and W.A. Bickmore. 2002b. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J Cell Biol.* 157:579-89.

Makarov, E.M., O.V. Makarova, H. Urlaub, M. Gentzel, C.L. Will, M. Wilm, and R. Luhrmann. 2002. Small Nuclear Ribonucleoprotein Remodeling During Catalytic Activation of the Spliceosome. *Science.*

Malatesta, M., C. Zancanaro, T.E. Martin, E.K. Chan, F. Amalric, R. Luhrmann, P. Vogel, and S. Fakan. 1994. Is the coiled body involved in nucleolar functions? *Exp Cell Res.* 211:415-9.

Mann, M., S.E. Ong, M. Gronborg, H. Steen, O.N. Jensen, and A. Pandey. 2002. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* 20:261-8.

Mann, M., and M. Wilm. 1995. Electrospray mass spectrometry for protein characterization. *Trends Biochem Sci.* 20:219-24.

Marciniak, R.A., D.B. Lombard, F.B. Johnson, and L. Guarente. 1998. Nucleolar localization of the Wemer syndrome protein in human cells. *Proc Natl Acad Sci U S A.* 95:6887-92.

Matera, A.G. 1999. Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol.* 9:302-9.

Matz, M.V., A.F. Fradkov, Y.A. Labas, A.P. Savitsky, A.G. Zaraisky, M.L. Markelov, and S.A. Lukyanov. 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol.* 17:969-73.

Maxwell, E.S., and M.J. Fournier. 1995. The small nucleolar RNAs. *Annu Rev Biochem.* 64:897-934.

McClintock, B. 1934. The relation of a particular chromosomal element to the development of the nucleoli in Zea mays. *Z Zellforsch Mikrosk Anat.* 21:294-328.

McNally, J.G., T. Karpova, J. Cooper, and J.A. Conchello. 1999. Three-dimensional imaging by deconvolution microscopy. *Methods.* 19:373-85.

Miller, O., and R. Beatty. 1969. Visualisation of nucleolar genes. *Science.* 164:955-57.

Misteli, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. *Science.* 291:843-7.

Misteli, T., and D.L. Spector. 1997. Applications of the green fluorescent protein in cell biology and biotechnology. *Nat Biotechnol.* 15:961-4.

Moneron, A., and W. Bernhard. 1969. Fine structural organization of the interphase nucleus in some mammalian cells. *J. Ultrastruct. Res.* 27:266-288.

Montgomery, T.H. 1898. Comparative cytological studies, with especial regard to the morphology of the nuclues. *J Morph.* 15:265-582.

Moss, T., and V.Y. Stefanovsky. 2002. At the center of eukaryotic life. *Cell.* 109:545-8.

Muratani, M., D. Gerlich, S.M. Janicki, M. Gebhard, R. Eils, and D.L. Spector. 2001. Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus. *Nat Cell Biol.* 21:21.

Nagorni, M., and S.W. Hell. 1998. 4Pi-confocal microscopy provides three-dimensional images of the microtubule network with 100- to 150-nm resolution. *J Struct Biol.* 123:236-47.

Narayanan, A., A. Lukowiak, B.E. Jady, F. Dragon, T. Kiss, R.M. Terns, and M.P. Terns. 1999a. Nucleolar localization signals of box H/ACA small nucleolar RNAs. *Embo J.* 18:5120-30.

Narayanan, A., W. Speckmann, R. Terns, and M.P. Terns. 1999b. Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell.* 10:2131-47.

Neubauer, G., A. King, J. Rappsilber, C. Calvio, M. Watson, P. Ajuh, J. Sleeman, A. Lamond, and M. Mann. 1998a. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet.* 20:46-50.

Neubauer, G., A. King, J. Rappsilber, C. Calvio, M. Watson, P. Ajuh, J. Sleeman, A. Lamond, and M. Mann. 1998b. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet.* 20:46-50.

Nigg, E.A. 1997. Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature.* 386:779-87.

Nissan, T.A., J. Bassler, E. Petfalski, D. Tollervey, and E. Hurt. 2002. 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. *Embo J.* 21:5539-47.

Nomura, M. 1999. Regulation of ribosome biosynthesis in Escherichia coli and Saccharomyces cerevisiae: diversity and common principles. *J Bacteriol.* 181:6857-64.

Ochs, R.L., T.W. Stein, Jr., and E.M. Tan. 1994. Coiled bodies in the nucleolus of breast cancer cells. *J Cell Sci.* 107:385-99.

Ogg, S.C., and A.I. Lamond. 2002. Cajal bodies and coilin--moving towards function. *J Cell Biol.* 159:17-21.

Olson, M.O., M. Dundr, and A. Szebeni. 2000. The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol.* 10:189-96.

Ong, S.E., B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, and M. Mann. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics.* 1:376-86.

Pandey, A., and M. Mann. 2000. Proteomics to study genes and genomes. *Nature.* 405:837-46.

Parada, L., and T. Misteli. 2002. Chromosome positioning in the interphase nucleus. *Trends Cell Biol.* 12:425-32.

Pasa-Tolic, L., R. Harkewicz, G.A. Anderson, N. Tolic, Y. Shen, R. Zhao, B. Thrall, C. Masselon, and R.D. Smith. 2002. Increased proteome coverage for quantitative peptide abundance measurements based upon high performance separations and DREAMS FTICR mass spectrometry. *J Am Soc Mass Spectrom.* 13:954-63.

Patterson, G.H., and J. Lippincott-Schwartz. 2002. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science.* 297:1873-7.

Pawley, J. 1995. Handbook of Biological Confocal Microscopy. Plenum Press.

Pederson, T. 1998. The plurifunctional nucleolus. *Nucleic Acids Res.* 26:3871-6.

Pederson, T. 1999. Movement and localization of RNA in the cell nucleus. *Faseb J.* 13 Suppl 2:S238-42.

Pederson, T. 2002. Dynamics and genome-centricity of interchromatin domains in the nucleus. *Nat Cell Biol.* 4:E287-91.

Pederson, T., and A. Kumar. 1971. Relationship between protein synthesis and ribosomal assembly in HeLa cells. *J Mol Biol.* 61:655-668.

Pederson, T., and J.C. Politz. 2000. The nucleolus and the four ribonucleoproteins of translation. *J Cell Biol.* 148:1091-5.

Phair, R.D., and T. Misteli. 2001. Kinetic modelling approaches to in vivo imaging. *Nat Rev Mol Cell Biol.* 2:898-907.

Platani, M., I. Goldberg, A.I. Lamond, and J.R. Swedlow. 2002. Cajal body dynamics and association with chromatin are ATP-dependent. *Nat Cell Biol.* 4:502-8.

Platani, M., I. Goldberg, J.R. Swedlow, and A.I. Lamond. 2000. In vivo analysis of Cajal body movement, separation, and joining in live human cells. *J Cell Biol.* 151:1561-74.

Pluta, A.F., A.M. Mackay, A.M. Ainsztein, I.G. Goldberg, and W.C. Earnshaw. 1995. The centromere: hub of chromosomal activities. *Science.* 270:1591-4.

Politz, J.C., L.B. Lewandowski, and T. Pederson. 2002. Signal recognition particle RNA localization within the nucleolus differs from the classical sites of ribosome synthesis. *J Cell Biol.* 159:411-8.

Politz, J.C., S. Yarovoi, S.M. Kilroy, K. Gowda, C. Zwieb, and T. Pederson. 2000. Signal recognition particle components in the nucleolus. *Proc Natl Acad Sci U S A.* 97:55-60.

Rabut, G., and J. Ellenberg. 2001. Nucleocytoplasmic transport: diffusion channel or phase transition? *Curr Biol.* 11:R551-4.

Rappsilber, J., U. Ryder, A.I. Lamond, and M. Mann. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12:1231-45.

Reits, E.A., and J.J. Neefjes. 2001. From fixed to FRAP: measuring protein mobility and activity in living cells. *Nat Cell Biol.* 3:E145-7.

Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol.* 17:1030-2.

Ritossa, F., and S. Spiegelman. 1965. Localization of DNA complementary to ribosomal RNA in the nucleolus organizer region of Drosophila melanogaster. *Proc Natl Acad Sci U S A.* 53:737-45.

Roepstorff, P. 1997. Mass spectrometry in protein studies from genome to function. *Curr Opin Biotechnol.* 8:6-13.

Roussel, P., C. Andre, C. Masson, G. Geraud, and D. Hernandez-Verdun. 1993. Localization of the RNA polymerase I transcription factor hUBF during the cell cycle. *J Cell Sci.* 104 (Pt 2):327-37.

Rout, M.P., G. Blobel, and J.D. Aitchison. 1997. A distinct nuclear import pathway used by ribosomal proteins. *Cell.* 89:715-25.

San-Segundo, P.A., and G.S. Roeder. 1999. Pch2 links chromatin silencing to meiotic checkpoint control. *Cell.* 97:313-24.

Santoro, R., J. Li, and I. Grummt. 2002. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet.* 32:393-6.

Scheer, U., and R. Hock. 1999. Structure and function of the nucleolus. *Curr Opin Cell Biol.* 11:385-90.

Scheer, U., B. Hugle, R. Hazan, and K.M. Rose. 1984. Drug-induced dispersal of transcribed rRNA genes and transcriptional products: immunolocalization and silver staining of different nucleolar components in rat cells treated with 5,6-dichloro-beta-D-ribofuranosylbenzimidazole. *J Cell Biol.* 99:672-9.

Scheer, U., and K.M. Rose. 1984. Localization of RNA polymerase I in interphase cells and mitotic chromosomes by light and electron microscopic immunocytochemistry. *Proc Natl Acad Sci U S A.* 81:1431-5.

Schul, W., L. de Jong, and R. van Driel. 1998. Nuclear neighbours: the spatial and functional organization of genes and nuclear domains. *J Cell Biochem.* 70:159-71.

Shaw, P.J., and E.G. Jordan. 1995. The nucleolus. *Annu Rev Cell Dev Biol.* 11:93-121.

Sherr, C.J., and J.D. Weber. 2000. The ARF/p53 pathway. *Curr Opin Genet Dev.* 10:94-9.

Shevchenko, A., D. Schaft, A. Roguev, W.W. Pijnappel, and A.F. Stewart. 2002. Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: analytical perspective. *Mol Cell Proteomics.* 1:204-12.

Shou, W., K.M. Sakamoto, J. Keener, K.W. Morimoto, E.E. Traverso, R. Azzam, G.J. Hoppe, R.M. Feldman, J. DeModena, D. Moazed, H. Charbonneau, M. Nomura, and R.J. Deshaies. 2001. Net1 stimulates RNA polymerase I transcription and regulates nucleolar structure independently of controlling mitotic exit. *Mol Cell.* 8:45-55.

Sinclair, D.A., and L. Guarente. 1997. Extrachromosomal rDNA circles--a cause of aging in yeast. *Cell.* 91:1033-42.

Sinclair, D.A., K. Mills, and L. Guarente. 1997. Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants. *Science.* 277:1313-6.

Singer, R.H., and M.R. Green. 1997. Compartmentalization of eukaryotic gene expression: causes and effects. *Cell.* 91:291-4.

Sirri, V., D. Hernandez-Verdun, and P. Roussel. 2002. Cyclin-dependent kinases govern formation and maintenance of the nucleolus. *J Cell Biol.* 156:969-81.

Sirri, V., P. Roussel, and D. Hernandez-Verdun. 2000. In vivo release of mitotic silencing of ribosomal gene transcription does not give rise to precursor ribosomal RNA processing. *J Cell Biol.* 148:259-70.

Sleeman, J., C.E. Lyon, M. Platani, J.P. Kreivi, and A.I. Lamond. 1998. Dynamic interactions between splicing snRNPs, coiled bodies and nucleoli revealed using snRNP protein fusions to the green fluorescent protein. *Exp Cell Res.* 243:290-304.

Sleeman, J.E., P. Ajuh, and A.I. Lamond. 2001. snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. *J Cell Sci.* 114:4407-19.

Sleeman, J.E., and A.I. Lamond. 1999. Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol.* 9:1065-74.

Smith, R. 2002. Trends in mass spectrometry instrumentation for proteomics. *Trends in Biotechnology.* 20:S3-S7.

Smith, R.D., G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, and H.R. Udseth. 2002. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics.* 2:513-23.

Snaar, S., K. Wiesmeijer, A.G. Jochemsen, H.J. Tanke, and R.W. Dirks. 2000. Mutational analysis of fibrillarin and its mobility in living human cells. *J Cell Biol.* 151:653-62.

Spector, D.L. 2001. Nuclear domains. *J Cell Sci.* 114:2891-3.

Stefanovsky, V.Y., G. Pelletier, R. Hannan, T. Gagnon-Kugler, L.I. Rothblum, and T. Moss. 2001. An immediate response of ribosomal transcription to growth factor stimulation in mammals is mediated by ERK phosphorylation of UBF. *Mol Cell.* 8:1063-73.

Straight, A.F., W. Shou, G.J. Dowd, C.W. Turck, R.J. Deshaies, A.D. Johnson, and D. Moazed. 1999. Net1, a Sir2-associated nucleolar protein required for rDNA silencing and nucleolar integrity. *Cell.* 97:245-56.

Strohner, R., A. Nemeth, P. Jansa, U. Hofmann-Rohrer, R. Santoro, G. Langst, and I. Grummt. 2001. NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. *Embo J.* 20:4892-900.

Sullivan, G.J., J.M. Bridger, A.P. Cuthbert, R.F. Newbold, W.A. Bickmore, and B. McStay. 2001. Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *Embo J.* 20:2867-74.

Sutherland, H.G., G.K. Mumford, K. Newton, L.V. Ford, R. Farrall, G. Dellaire, J.F. Caceres, and W.A. Bickmore. 2001. Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Hum Mol Genet.* 10:1995-2011.

Swedlow, J.R., K. Hu, P.D. Andrews, D.S. Roos, and J.M. Murray. 2002. Measuring tubulin content in Toxoplasma gondii: a comparison of laser-scanning confocal and wide-field fluorescence microscopy. *Proc Natl Acad Sci U S A.* 99:2014-9.

Swedlow, J.R., and A.I. Lamond. 2001. Nuclear dynamics: where genes are and how they got there. *Genome Biol.* 2:REVIEWS0002.

Swedlow, J.R., and M. Platani. 2002. Live cell imaging using wide-field microscopy and deconvolution. *Cell Struct Funct.* 27:335-41.

Terskikh, A., A. Fradkov, G. Ermakova, A. Zaraisky, P. Tan, A.V. Kajava, X. Zhao, S. Lukyanov, M. Matz, S. Kim, I. Weissman, and P. Siebert. 2000. "Fluorescent timer": protein that changes color with time. *Science.* 290:1585-8.

Thiry, M., T. Cheutin, M.F. O'Donohue, H. Kaplan, and D. Ploton. 2000. Dynamics and three-dimensional localization of ribosomal RNA within the nucleolus. *Rna.* 6:1750-61.

Thomann, D., D.R. Rines, P.K. Sorger, and G. Danuser. 2002. Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. *J Microsc.* 208:49-64.

Truong, K., and M. Ikura. 2001. The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Curr Opin Struct Biol.* 11:573-8.

Tsai, R.Y., and R.D. McKay. 2002. A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. *Genes Dev.* 16:2991-3003.

Tsien, R.Y. 1998. The green fluorescent protein. *Annu Rev Biochem.* 67:509-44.

Tsukamoto, T., N. Hashiguchi, S.M. Janicki, T. Tumbar, A.S. Belmont, and D.L. Spector. 2000. Visualization of gene activity in living cells. *Nat Cell Biol.* 2:871-8.

Vale, R.D. 2000. AAA proteins. Lords Of the ring. *J Cell Biol.* 150:F13-20.

Valentin, G. 1836. Repertorium fur Anatomie und Physiologie. Verlag von Veit und Comp, Berlin. 1-293 pp.

Van Craenenbroeck, E., and Y. Engelborghs. 2000. Fluorescence correlation spectroscopy: molecular recognition at the single molecule level. *J Mol Recognit.* 13:93-100.

Van Leeuwenhoek, A. 1948. Letter no. 35 of March 3, 1682. Vol. 67. Alle de Brieven van Antoni van Leeuwenhoek, Utigever Swets en Zeitlinger, Amsterdam. Deel III, 382.

van Roessel, P., and A.H. Brand. 2002. Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins. *Nat Cell Biol.* 4:E15-20.

Visintin, R., and A. Amon. 2000. The nucleolus: the magician's hat for cell cycle tricks. *Curr Opin Cell Biol.* 12:372-7.

Visintin, R., E.S. Hwang, and A. Amon. 1999. Cfi1 prevents premature exit from mitosis by anchoring Cdc14 phosphatase in the nucleolus. *Nature.* 398:818-23.

Voit, R., and I. Grummt. 2001. Phosphorylation of UBF at serine 388 is required for interaction with RNA polymerase I and activation of rDNA transcription. *Proc Natl Acad Sci U S A.* 98:13631-6.

Voit, R., M. Hoffmann, and I. Grummt. 1999. Phosphorylation by G1-specific cdk-cyclin complexes activates the nucleolar transcription factor UBF. *Embo J.* 18:1891-9.

Wall, M.A., M. Socolich, and R. Ranganathan. 2000. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat Struct Biol.* 7:1133-8.

Wanzenbock, E.M., C. Schofer, D. Schweizer, and A. Bachmair. 1997. Ribosomal transcription units integrated via T-DNA transformation associate with the nucleolus and do not require upstream repeat sequences for activity in Arabidopsis thaliana. *Plant J.* 11:1007-16.

Warner, J.R. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci.* 24:437-40.

Warner, J.R. 2001. Nascent ribosomes. *Cell.* 107:133-6.

Warner, J.R., and R. Soeiro. 1967. Nascent ribosomes from HeLa cells. *Proc Natl Acad Sci U S A.* 58:1984-1990.

Weinstein, L.B., and J.A. Steitz. 1999. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol.* 11:378-84.

Weis, K. 1998. Importins and exportins: how to get in and out of the nucleus. *Trends Biochem Sci.* 23:185-9.

Weis, K. 2002. Nucleocytoplasmic transport: cargo trafficking across the border. *Curr Opin Cell Biol.* 14:328-35.

Weisenberger, D., and U. Scheer. 1995. A possible mechanism for the inhibition of ribosomal RNA gene transcription during mitosis. *J Cell Biol.* 129:561-75.

Wilm, M., and M. Mann. 1996. Analytical properties of the nanoelectrospray ion source. *Anal Chem.* 68:1-8.

Wolffe, A.P., and J.C. Hansen. 2001. Nuclear visions: functional flexibility from structural instability. *Cell.* 104:631-4.

Wong, J.M., L. Kusdra, and K. Collins. 2002. Subnuclear shuttling of human telomerase induced by transformation and DNA damage. *Nat Cell Biol.* 4:731-6.

Yang, F., L.G. Moss, and G.N. Phillips, Jr. 1996. The molecular structure of green fluorescent protein. *Nat Biotechnol.* 14:1246-51.

Yang, Y., Y. Chen, C. Zhang, H. Huang, and S.M. Weissman. 2002. Nucleolar localization of hTERT protein is associated with telomerase function. *Exp Cell Res.* 277:201-9.

Yao, X., A. Freas, J. Ramirez, P.A. Demirev, and C. Fenselau. 2001. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem.* 73:2836-42.

Yarbrough, D., R.M. Wachter, K. Kallio, M.V. Matz, and S.J. Remington. 2001. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. *Proc Natl Acad Sci U S A.* 98:462-7.

Zatsepina, O.V., R. Voit, I. Grummt, H. Spring, M.V. Semenov, and M.F. Trendelenburg. 1993. The RNA polymerase I-specific transcription initiation factor UBF is associated with transcriptionally active and inactive ribosomal genes. *Chromosoma.* 102:599-611.

Zhou, Y., R. Santoro, and I. Grummt. 2002a. The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *Embo J.* 21:4632-40.

Zhou, Z., L.J. Licklider, S.P. Gygi, and R. Reed. 2002b. Comprehensive proteomic analysis of the human spliceosome. *Nature.* 419:182-5.

CHAPTER II

MATERIALS AND METHODS

"Unless the Lord builds the house,
its builders labour in vain."

~ Psalm, The Bible ~

II. Materials and Methods

II.1 Bacterial strains and general solutions

II.1.1 Bacterial strains

Epicurian coli® SoloPack™ supercompetent cells (Strategene #230350): Tet^R *Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte (F' *proAB lacI^{q}ZΔM15* Tn*10*(Tet^R) Amy Cam^R). This bacterial strain is very efficient in bacterial transformation especially for ligated product: (1) the absence of endonuclease (*endA1*) and recombinase (*recA*) improves the quality and stability of plasmid DNA miniprep, respectively; and (2) the lack of the restriction systems (*mcrA mcrCB and mrr*) and the mutation in the endonuclease *hsd*R gene facilitates the introduction of DNA isolated from non-*Escherichia coli* organism.

II.1.2 General solutions

50x TAE Buffer

242g	Tris base
57.1ml	Glacial acetic acid
100ml	0.5M EDTA (pH8.0)
1L	**Total**

10x TBE Buffer

84g	Tris base
55g	Boric acid
40ml	0.5M EDTA (pH8.0)
1L	**Total**

Luria-Bertani Medium (LB)

10g	Bacto-tryptone
5g	Bacto-yeast extract
10g	NaCl
950ml	*Subtotal* (with MQ H_2O)

pH to 7.0 using 5M NaOH (~0.2ml)
adjust the volume to 1L with MQ H_2O
Solution autoclaved

Terrific Broth (TB)

12g	Bacto-tryptone
24g	Bacto-yeast extract
4ml	Glycerol
900ml	*Subtotal* (with MQ H_2O)

Solution autoclaved

100ml	Sterile 0.17M KH_2PO_4
1L	**Total**

Phosphate Buffer Saline (PBS)

8g	NaCl
0.2g	KCl
1.44g	Na_2HPO_4
0.24g	KH_2PO_4
800ml	**Total** (with MQ H_2O)

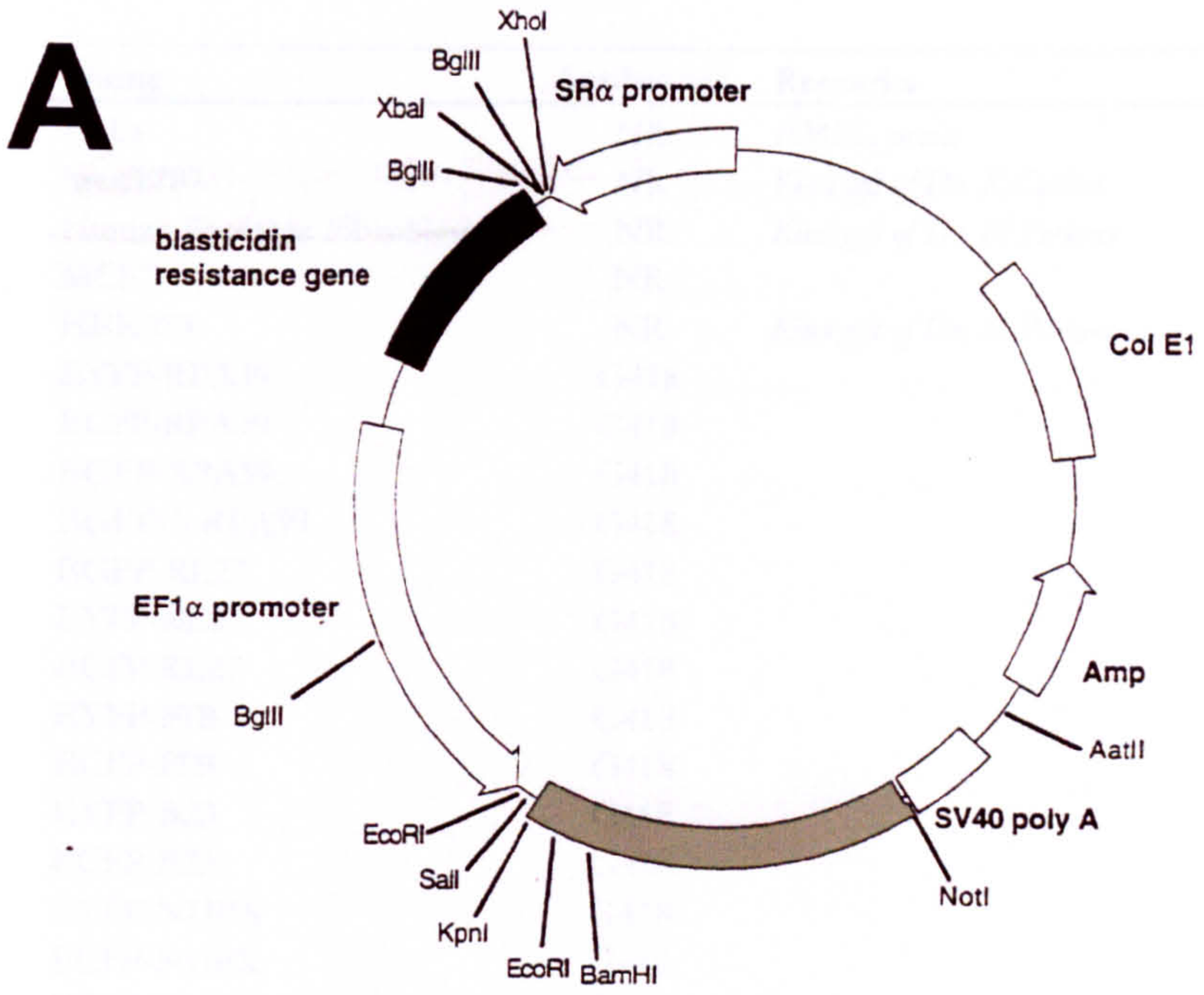
PBS-T

2.5ml	10% Tween-20 (w/v in PBS)
500ml	PBS

II.2 Plasmids generated and used

Name	Cloning sites	Backbone vector	Antibiotics	Remarks
EYFP-RPA39	BglII, KpnI	EYFP-C1	G418	
ECFP-RPA39	BglII, KpnI	ECFP-C1	G418	
EGFP-RPA39	BglII, KpnI	EGFP-C1	G418	
EGFPx3-RPA39	BglII, KpnI	EGFPx3-C1	G418	
EGFP-RL27	BglII/BamHI, EcoRI	EGFP-C1	G418	
EYFP-RL27	BglII/BamHI, EcoRI	EYFP-C1	G418	
ECFP-RL27	BglII/BamHI, EcoRI	ECFP-C1	G418	
DsRed1-RL27	BglII/BamHI, EcoRI	DsRed1-C1	G418	
DsRed2-RL27	EcoRI, BamHI	DsRed2-C1	G418	
EYFP-FIB	BglII, KpnI	EYFP-C1	G418	
ECFP-FIB	BglII, KpnI	ECFP-C1	G418	pAL118
DsRed1-FIB	BglII, KpnI	DsRed1-C1	G418	
Bos EYFP-FIB	NotI, BamHI	pBos	Blasticidin	
Bos ECFP-FIB	NotI, BamHI	pBos	Blasticidin	
Bos EYFP-B23	NotI, BamHI	pBos	Blasticidin	
EYFP-B23	KpnI, BamHI	EYFP-C1	G418	
ECFP-B23	KpnI, BamHI	ECFP-C1	G418	
DsRed1-B23	KpnI, BamHI	DsRed1-C1	G418	
DsRed2-B23	KpnI, BamHI	DsRed2-C1	G418	
EYFP-NHPX	BglII, KpnI	EYFP-C1	G418	pAL107
ECFP-NHPX	BglII, KpnI	ECFP-C1	G418	pAL214
EGFP-NHPX	BglII, KpnI	EGFP-C1	G418	
pTimer1-NHPX	BglII, KpnI	pTimer1-C1	G418	
H2B-EYFP	NotI, BamHI	pBos	Blasticidin	
H2B-EGFP	NotI, BamHI	pBos	Blasticidin	*Kind gift of Dr. T. Kanada*
H2B-ECFP	NotI, BamHI	pBos	Blasticidin	
diHcRed-H2B			G418	*Kind gift of Dr. J. Ellenberg*
EGFP-PWP1	EcoRI, BamHI	EGFP-C1	G418	
EYFP-PWP1	EcoRI, BamHI	EYFP-C1	G418	
EGFP-NGB	HinDIII, EcoRI	EGFP-C1	G418	
ECFP-NGB	HinDIII, EcoRI	ECFP-C1	G418	
EYFP-NGB	HinDIII, EcoRI	EYFP-C1	G418	
EYFP-2H9	KpnI, BamHI	EYFP-C1	G418	
ECFP-MrDB	KpnI, BamHI	ECFP-C1	G418	
EYFP-MrDB	KpnI, BamHI	EYFP-C1	G418	
EYFP-PES1	BglII, EcoRI	EYFP-C1	G418	
U1A-DsRed1	EcoRI, KpnI	DsRed1-N1	G418	*Kind gift of Dr. L. Trinkle-Mulcahy*
EYFP-PWP2	HinDIII, BamHI	EYFP-C1	G418	
EYFP-WDR3	BglII, KpnI	EYFP-C1	G418	
EYFP-PBK1	BglII/BamHI, EcoRI	EYFP-C1	G418	
EYFP-SAZD	BglII, KpnI	EYFP-C1	G418	

Table II-1 **Table of plasmids generated and used.** *'Cloning sites' indicates the restriction enzyme sites used for ligation to the 'backbone vector'.* 'Antibiotics' indicates the drugs that can be used for generating a stable cell line. The vector maps are shown in Figure II-1. Apart from those indicated, all other plasmids were generated during the PhD and other plasmids that were not mentioned in this thesis but generated during the PhD were deposited to the Lamond lab reagent database. Backbone vector EGFPx3-C1 is a kind gift from Dr. J. Ellenberg.



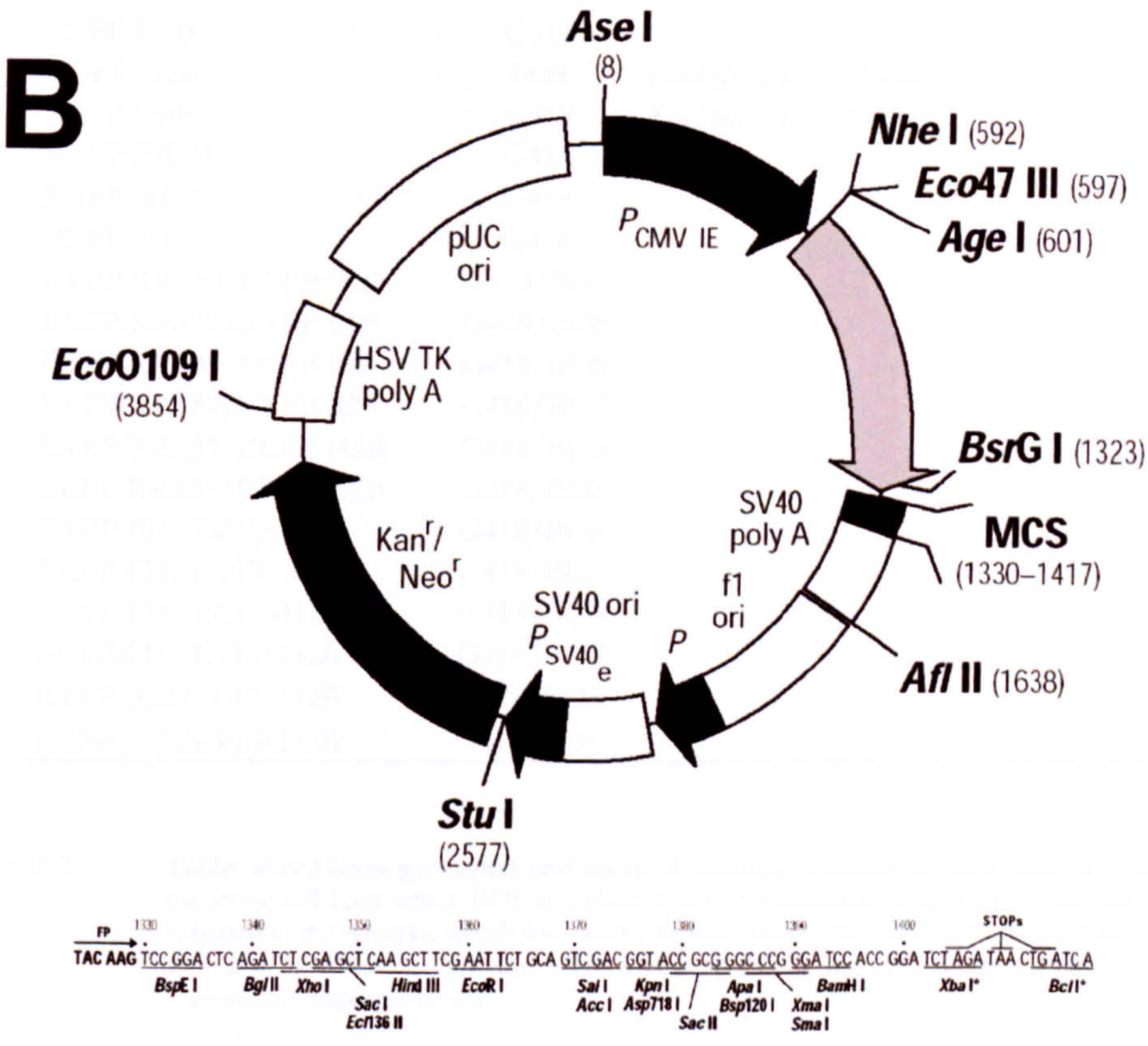


Figure II-1 **Plasmid maps.** (A) pBOS (6kbp) and (B) a general scheme for pFP-C1 vectors (4.7kbp) including EYFP-C1, ECFP-C1, EGFP-C1, EGFPx3-C1, DsRed1-C1, DsRed2-C1 (Clontech). Grey areas indicate the location of the fluorescent protein fusion. The multiple cloning sites for the pFP-C1 vector are shown below the map.

II.3 Cell lines generated and used

Name	Antibiotics	Remarks
HeLa	NR	*EMBL strain*
htert1787	NR	*Kind gift of Dr. K Collins*
Human Foreskin Fibroblast	NR	*Kind gift of Dr. N Perkins*
MCF7	NR	
HEK293	NR	*Kind gift of Dr. N Perkins*
EYFP-RPA39	G418	
ECFP-RPA39	G418	
EGFP-RPA39	G418	
EGFPx3-RPA39	G418	
EGFP-RL27	G418	
EYFP-RL27	G418	
ECFP-RL27	G418	
EYFP-FIB	G418	
ECFP-FIB	G418	
EYFP-B23	G418	
ECFP-B23	G418	
EYFP-NHPX	G418	
ECFP-NHPX	G418	
EGFP-NHPX	G418	
EYFP-H2B	G418	
EGFP-H2B	BOS	*Kind gift of Dr. T Kanada*
ECFP-SmB	G418	*Kind gift of Dr. J Sleeman*
EYFP-PWP1	G418	
EYFP-NGB	G418	
EYFP-2H9	G418	
EYFP-RPA39/ECFP-FIB	G418/BOS	
ECFP-RPA39/EYFP-FIB	G418/BOS	
EYFP-RPA39/ECFP-B23	G418/BOS	
ECFP-RPA39/EYFP-B23	G418/BOS	
EYFP-RPA39/ECFP-H2B	G418/BOS	
ECFP-RPA39/EYFP-H2B	G418/BOS	
EYFP-FIB/ECFP-B23	G418/BOS	
ECFP-FIB/EYFP-B23	G418/BOS	
EYFP-FIB/ECFP-H2B	G418/BOS	
ECFP-FIB/EYFP-H2B	G418/BOS	
EYFP-B23/ECFP-H2B	G418/BOS	
ECFP-B23/EYFP-H2B	G418/BOS	

Table II-2 **Table of cell lines generated and used.** '*Antibiotics*' indicates the drug used to generate the stable cell lines where *BOS* is a short form of blasticidin. Apart from those cell lines indicated in the remarks, all cell lines were generated during the PhD and other cell lines that were not mentioned in this thesis but generated during the PhD were deposited in the Lamond lab reagent database.

II.4 Molecular Biology Techniques

II.4.1 "3-day" cloning strategies

To facilitate the high throughput screening of potential components of nucleoli from the proteomic study, I have devised a cloning strategy that will allow multiple cloning to be performed within 72 hours. The scheme of the strategy is illustrated in Figure II-2 and the protocol for each procedure will be explained in detail in the following sections (II.4.2-II.4.11)

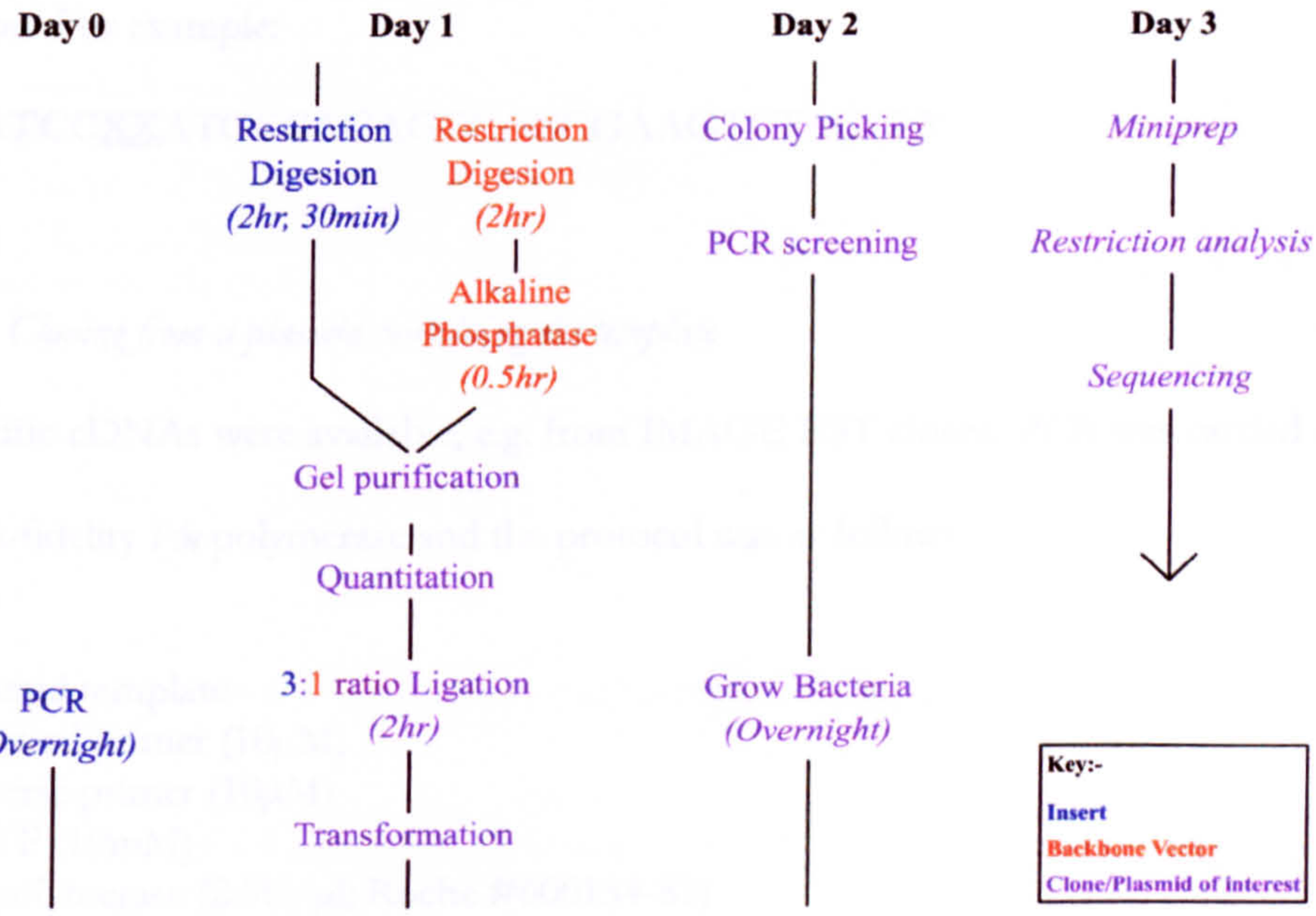


Figure II-2 **Scheme of the "3-day" cloning strategies**

II.4.2 Polymerase Chain Reaction (PCR)

II.4.2.1 General

To amplify the DNA fragments of interest to generate a C-terminal fusion protein as in most of the work presented in this study, two specific primers hybridising to the 5' and 3' end of the complementary DNA (cDNA) were designed with the following criteria: (1) G/C content of at least 50%; (2) the last 3' nucleotide is either G or C but the last three

nucleotides must not be the same; (3) length of about 24bp which has a perfect homology with the gene of interest; (4) the restriction site hanger (indicated in **bold** below) is located 5' to the length of perfect homology with spacers (indicated with an underline below) to make sure that the protein of interest would be in frame with the N-terminal fusion and (5) extra nucleotides (indicated as *italic* below) were added 5' to the restriction site hanger to ensure optimal restriction digestion after the PCR. The number of nucleotides in *italics* depended on the enzymes in the restriction site hanger according to the manufacturer's suggestion. Moreover, the sequences are the inverse of the 3' end to avoid any possible hybridization. For example:

5'-*GG***GGATCCXX**ATGCCAGAGCCAGCGAAGTCTAGG-3'

II.4.2.2 Cloning from a plasmid containing the template

Where specific cDNAs were available, e.g. from IMAGE EST clones, PCR was carried out using a high-fidelity *Pfu* polymerase and the protocol was as follows:

Materials:

1-2ng	plasmid template
1µl	Forward primer (10µM)
1µl	Reverse primer (10µM)
1µl	dNTP (10mM)
1µl	*pfu* polymerase (2.5U/µl; Roche #600154-81)
5µl	10x polymerase buffer
50µl	**Total** (with MQ H_2O)

Cycle:

Pre-PCR

95°C	5 min

PCR 25 cycles

95°C	1 min
55-65°C	30 sec
72°C	x min

(where x= length of the coding sequence ÷ 1kbp/min)

Post-PCR

72°C	x min *(where x= length of the coding sequence ÷ 1kbp/min)*
4°C	∞ *until use or kept at -20°C*

The PCR product was purified using the QIAquick PCR purification kit and eluted in a volume of 30µl MQ H_2O (QIAGEN #28104).

II.4.2.3 Cloning from a cDNA library

For those genes of interest that could not be cloned from an available EST source (e.g. IMAGE), the cDNA was fished out from a pre-made double stranded Marathon-Ready™ cDNA library (Clontech #K1802-1) as follows:

Materials:

5µl	Marathon cDNA library (~0.1ng/µl)
1µl	Forward primer (10µM)
1µl	Reverse primer (10µM)
1µl	dNTP (10mM)
1µl	Advantage 2 Polymerase (Clontech #8430-1)
5µl	10x polymerase buffer
50µl	**Total** (with MQ H_2O)

Cycle:

Pre-PCR

95°C	5 min

PCR1 25 cycles

95°C	1 min
68°C	x min *(where x= length of the coding sequence ÷ 1kbp/min)*

Post-PCR

68°C	x min *(where x= length of the coding sequence ÷ 1kbp/min)*
15°C	∞ *until use or kept at -20°C*

The PCR product was diluted in 500-1000 fold and 5µl of the dilution was subjected to another round of PCR using *pfu* polymerase as described in Section II.4.2.2.

II.4.2.4 PCR screening

To choose the right clone from the bacterial transformation (Section II.4.8), screening was initially carried out by PCR. A set of PCR primers was chosen such that one primer is hybridising to the backbone vector, whilst the other primer hybridises to the insert at the opposite end. In this way, both the orientation and size of the fragment cloned into the

vector could be selected. Another advantage of PCR screening is that bacterial colonies can be used as the starting material because the plasmid DNA is released during the first PCR step at 95°C for 5 minutes. This could save time and materials in producing unwanted minipreps. However, great care should be made in handling different bacteria clones to avoid cross-contamination. The following master mix was used:

Materials:

1	Bacterial clone
0.5µl	Forward vector-specific primer (10µM)
0.5µl	Reverse insert-specific primer (10µM)
0.5µl	dNTP (10mM)
0.25µl	Advantage 2 Polymerase (Clontech #8430-1)
2.5µl	10x polymerase buffer
25µl	**Total** (with MQ H_2O)

Cycle:

Pre-PCR

95°C	5 min

PCR1 10 cycles

95°C	1 min
68°C	x min *(where x= length of the coding sequence ÷ 1kbp/min)*

PCR1 20 cycles

95°C	5 sec
68°C	x min *(where x= length of the coding sequence ÷ 1kbp/min)*

Post-PCR

68°C	x min *(where x= length of the coding sequence ÷ 0.5kbp/min)*
15°C	∞ *until use or kept at -20°C*

10µl of resultant mixture was mixed with loading dye and visualized on the agarose gel as described in Section II.4.5.

II.4.3 Restriction Digestion

II.4.3.1 General

For a digestion involving a single enzyme, the reaction was performed as suggested by the manufacturer's suggestion. However, for a digestion involving two enzymes, the reaction was carried out in a buffer such that both enzymes have at least 75% activity. However, if

such a buffer is not available or the enzymes involved are optimal for different temperatures (e.g. SmaI is active at 25°C whereas most of the other enzymes work best at 37°C), a sequential digestion was performed first with the buffer of lower salt concentration. To avoid star activity of some enzymes, according to the manufacturer's suggestions, BSA was added at a final concentration of 100μg/ml.

II.4.3.2 Digestion for subsequent ligation reactions

Unless a digested backbone vector was available, the digestion of insert and backbone vector was carried out in parallel as described below:

Backbone Vector

2μg	Backbone vector
10U	restriction enzyme 1
[10U	restriction enzyme 2]
[0.3μl	BSA (10mg/ml)]
3μl	10x enzyme buffer
30μl	**Total** (with MQ H_2O)

Insert

10μg	Insert
10U	restriction enzyme 1
[10U	restriction enzyme 2]
[0.3μl	BSA (10mg/ml)]
3μl	10x enzyme buffer
30μl	**Total** (with MQ H_2O)

The reactions of the backbone vector and insert were carried out at 37°C, unless specified, for 2.5 hours. For the last half hour in the backbone vector reaction, alkaline phosphatase was added to remove the 3' phosphate group to increase ligation efficiency (see Section II.4.4 for details).

II.4.3.3 Digestion for restriction analysis

Restriction analysis was carried out to further confirm the orientation if the results of PCR screening (Section II.4.2.3) were not satisfactory. Protocol as follows:

Restriction analysis

400ng	Miniprep
10U	restriction enzyme 1
[10U	restriction enzyme 2]
[0.2µl	BSA (10mg/ml)]
2µl	10x enzyme buffer
20µl	**Total** (with MQ H_2O)

The reaction was carried out at 37°C, unless specified, for 1 hour. The resulting mixtures were loaded onto 0.8%-1.5% agarose gel depending on the expected size of the DNA fragment (see Section II.4.5 for details).

II.4.4 Alkaline Phosphatase treatment of Plasmid DNA

To avoid the self-ligation of the vector, if both ends are compatible, 3' phosphate groups were removed from the linearised vector by treating with alkaline phosphatase. 1µl of alkaline phosphatase (Roche #1097075) was added to the backbone vector reaction in the last half hour (Section II.4.3.2). And the resultant mixture was immediately run on a 0.8%-1.5% agarose gel, depending on the expected size of the fragments, for purification (see Section II.4.5 for details).

II.4.5 Gel Electrophoresis of DNA

II.4.5.1 Loading dye

The DNA loading dye was prepared as follows *(Sambrook et al.)*:

6x DNA Loading Dye

0.25%	bromophenol blue
0.25%	xylene cyanol FF
30%	glycerol

II.4.5.2 Preparation of agaorse gel and running samples

To prepare an agarose gel, depending on the percentage of agarose required (0.8-1.5%), the correct amount of powdered agarose (BDH #443665W) was added to a measured quantity of 1x TAE buffer (Section II.1) in a glass flask. The slurry was heated in a microwave at

low power until dissolved without reaching the boiling point. The mixture was cooled to about 60°C before adding Ethidium Bromide (EtBr) to a final concentration of about 20ng/ml). The cooled mixture was poured into the mould making appropriate number of wells. The set gel was then submerged into 1x TAE buffer for loading and electrophoresis. The DNA samples after mixing with the loading dye were loaded onto the wells with one or more lanes dedicated for markers (Bioline #BIO-33026). Electric field was then applied at a final voltage of 80-100V for roughly 2 hour, depending on the expected size of the fragments.

II.4.5.3 Purification of DNA from agarose gels

To recover the DNA fragments from agarose gel for ligation, the gel was illuminated by a light box emitting a filtered blue light (Clare Chemical Research Dark Reader™ DR-45M) and each fragment was cut out from the gel using a clean scalpel. The excised gel was weighed and DNA was extracted with 30µl MQ H_2O using QIAquick gel extraction kit (QIAGEN #28704).

II.4.5.4 Preparation of "mini" DNA gel for quantitation

To quantify the amount of fragments while saving most of the materials for subsequent ligation, a "mini" gel was made in a way similar to Section II.4.5.2. Rather than pouring into a large mould, 25ml cooled slurry (without EtBr) was poured onto an ethanol-cleaned 6"x4" glass plate with a small-teeth comb clamped perpendicularly, allowing the surface tension to spread the slurry evenly on top of the glass.

II.4.6 Quantitation of plasmid : insert ratio

The set mini-gel (Section II.4.5.4) was then submerged in 1x TAE buffer and 1µl of the resultant fragment DNA was loaded with dye onto the tiny wells. The gel was run at 80V

for 1 hour before staining by a highly sensitive DNA dye SYBR Green I (Molecular Probes #S-7563) for 15 minutes in the dark. The intensity of the DNA fragment was measured by Fujifilm Intelligent Dark Box LAS-1000 with a Fluorescence filter Y515-Di.

II.4.7 Ligation

From the quantitation data, the intensity of each band was first subtracted from background and then normalized by the corresponding fragment size. A molar ratio of 3:1 for the amount of insert and backbone vectors is required for a ligation reaction.

Ligation Reaction

50ng	linearised vector
xµl	insert *(xµl was added such that the insert : backbone vector molar ratio=3:1)*
0.5µl	ligase (1 U/µl, Roche #716359)
[1µl	10mM ATP *for cohesive end ligation*]
1µl	10x ligase buffer
10µl	**Total** (with MQ H_2O)

The ligation reaction was carried at 12-16°C for 2 hours.

II.4.8 Bacterial Transformation

After the ligation, 1µl of the resultant mixture was transformed into Epicurian Coli®SoloPack™ Gold supercompetent cells (Strategene #230350, Section II.1.1) according to the manufacturer's suggestion and the transformed cells were plated on to LB plate containing the selected antibiotics:

Antibiotics used

Carbenicillin (Ampicillin)	100µg/ml (f/c)
Kanamycin	30µg/ml (f/c)

II.4.9 Preparation of plasmid DNA

II.4.9.1 Day 3 for "3-day" cloning strategies

On Day 3 (Section II.4.1), single colonies were picked individually using a P-200 Gilson pipette tip for 3 different purposes: (1) PCR screening (Section II.4.2.3); (2) temporary

storage: by streaking selected clones on a new LB plate with appropriate antibiotics and (3) mini-preparation of the plasmid. The bacteria-containing tips were inoculated in 3ml LB with appropriate antibiotics in 15ml sterile plastic tubes (Falcon #352059) and the tube were shaken at 250rpm at 37°C for 12-16 hours. The culture was harvested by QIAGEN Plasmid miniprep kit according to the manufacturer's suggestion (QIAGEN #27106), which is based on a modified alkaline lysis procedure (Sambrook et al.). The resultant plasmid was eluted with 50μl MQ H_2O and stored at -20°C until use.

II.4.9.2 Large scale preparation of plasmid DNA

A bacterial clone of interest was picked from the storing plate as mentioned in Section II.4.9.1 and was allowed to grow in 100ml TB (Section II.1.2). After 12-16 hour incubation with shaking (300rpm) at 37°C, the plasmid was then harvested by QIAGEN Plasmid maxiprep kit according to the manufacturer's suggestion (QIAGEN #12163). The resultant plasmid was eluted with 500-1000μl MQ H_2O, depending on the size of the corresponding bacterial pellet and stored in aliquots at -20°C until use.

II.4.10 Spectrophotometric determination of DNA concentration

To determine the DNA concentration, the plasmid stock was diluted 100 fold with MQ H_2O and measured by Gene QuantII (Pharmacia Biotech) in triplicate. Absorbance at 260nm and 280nm were taken to measure the DNA and protein concentration, respectively, in the sample, where a pure sample should have the ratio of A_{260} /A_{280} between 1.8 and 2.0, using a quartz curvette (Hellma; light path=10mm)

II.4.11 DNA sequencing

The sequence of the insert at the junctions with the vector backbone was initially confirmed by using primers hybridized to both ends of the backbone vector. The internal

region, if required, was confirmed by designing primers with the following criteria: (1) ~50% GC content; (2) length of about 17bp and (3) ends with G/C but both the first three and last three nucleotides should not be identical. The DNA sequencing was performed using ABI Prism Big Dye Terminator cycle sequencing reaction by DNA sequencing service (http://www.dnaseq.co.uk/)

Materials submitted

300ng	plasmid
1µl	specific primer (3.2µM)
10µl	**Total** (with MQ H_2O)

The results were analysed using program Sequencing Manager (Lasergene) to assemble a single contig and the resultant contig was checked both at the junction and with the known coding sequence. The latter sequence check was performed using the program BLAST2 http://www.ncbi.nlm.nih.gov/BLAST/ (Altschul et al., 1990) and the locations of ambiguity were further examined on the sequencing profiles or by performing more sequencing using primers that hybridised to the alternate strand, if needed.

II.5 Tissue Culture

II.5.1 Basic cell culture protocols

II.5.1.1 Medium preparation

In most of the cell lines, such as HeLa, MCF7 and HEK293 etc, that were grown in the laboratory for the purpose of this thesis, the following was used unless specified.

Standard medium

50ml	Fetal Bovine Serum (FCS; Invitrogen #10106-169)
5ml	Penicillin-Streptomycin (PS; Invitrogen #15140-122)
445ml	Dulbecco's Modified Eagle Medium (DMEM; Invitrogen #41966-052)
500ml	**Total**

For most of the HeLa cell lines stably expressing one or more fluorescent fusion proteins, the following selectable markers were used at the following concentrations after filter-sterilising (See Section II.3 for details):

Specific antibiotics used for generating cell lines

G418	200µg/ml (f/c; Roche #146-4981)
Blasticidin	2µg/ml (f/c; Invitrogen #25-0205)

II.5.1.2 General growth conditions

The cell lines were grown in 75cm^2 flasks with filter caps (Greiner #658175) in a humidified incubator set at 5% CO_2 and 37°C. When the cells reached 70-100% confluency, the cells were split using 1ml Trypsin (Invitrogen #25300-054) per flask for 4 minutes and neutralised with the standard medium. Subsequently, the cells were diluted 1 in 10 for replating. All tissue culture work was performed inside a laminar flow hood. All medium and apparatus used were first sprayed with 70% ethanol before putting into the hood to avoid contamination.

II.5.2 Transfection

Two different transfection procedures were commonly preformed in this study, depending on the cell types. HEK293 cells are best transfected with Calcium Phosphate Precipitation procedure while other cell lines are mostly transfected with Effectene Reagent (QIAGEN #301425).

II.5.2.1 Calcium Phosphate Precipitation Transfection

Cells were split onto a 90mm dish (Merck #402/0322/18) or 140mm dish (Merck #402/0322/20) to ~20-30% confluency at least 2 hours prior to transfection, allowing enough time for the cells to adhere. 2x HBS was thawed at room temperature and the transfection mixture prepared as follows:

DNA/$CaCl_2$ mix

10µg	DNA
61µl	2M $CaCl_2$
500µl	**Total** *(with MQ H_2O)*

2x HBS

8mg NaCl
0.2mg $Na_2HPO_4.7H_2O$ *([phosphate]=1.5mM)*
6.5mg HEPES
500µl Total (with MQ H_2O and pH7.0)

All solutions used were filter-sterilised. To make the transfection precipitate, DNA/$CaCl_2$ mix were added to the 2x HBS in a 15 ml falcon tube (WTS #TPP15) dropwise using a P-1000 Gilson pipette and mixed gently by slight tapping during the addition. The mixture was then added immediately to the cells slowly and evenly into the medium. Swirling should be avoided at this stage and the precipitate should have been mixed when the dishes were transferred back to the humidified incubator.

II.5.2.2 Using Effectene

The Effectene reagent is a non-liposomal lipid and was always used in conjunction with the Enhancer and the DNA-condensation buffer to achieve high transfection efficiency. The method is generally based on the manufacturer's suggestion and usually 1µg of plasmid was transfected per 60mm dish (Merck #402/0322/12) and 2µg per 90mm (Merck #402/0322/18). All of the cell lines generated in this study used this reagent (Section II.5.5).

II.5.3 Microinjection

To perform the microinjection of plasmids into cultured cells, high quality DNA is required and hence a large-scale, high quality plasmid preparation was made according to Section II.4.9.2. The concentration was diluted to 20µg/ml for optimal expression using injection buffer as described below. In some experiments, 100-500ng/ml TexasRed fixable dextran (Sigma) was added to help locate microinjected cells. The diluted DNA was then spun at 14,000 rpm in a bench centrifuge for 15min at 4°C (Eppendorf 5415c) to avoid any

microparticulate matter that may block the injection needle (Eppendorf sterile Femtotips). The microinjection was performed using the microinjector (Eppendorf 5246) on a Zeiss Axiovert S100 on a heated stage at 37°C. The cells were seeded one day earlier on a grid coverslip (Eppendorf Celllocate® square size 175µm) in a 60mm dish (Merck #402/0322/12). The cells of interest were initially focused by a 10x lens (Zeiss Plan-NeoFluar 0, 30) followed by a 40x lens (LD Achroplan 0, 60, Korr Ph2) for fine adjustment. The microinjection was performed on about 300 cells/coverslip at a pressure between 80-110Pa.

Injection Buffer

14.7g	Glutamic acid
7.85g	KOH
1ml	1M $MgSO_4$
1ml	1M DTT
1L	**Total** *(with MQ H_2O)*

pH to 7.2 with 1 M citric acid

II.5.4 Heterokaryon formation

Two different cell lines were mixed in a 1:1 ratio at a total of 50% confluency onto 90mm dish (Merck #402/0322/18) one day before the experiment. When the cells reached around 90-100% confluency, just enough polyethylene glycol (PEG; SIGMA #P7181) to cover the cells was added and the whole dish was rocked for 90 seconds to ensure thorough mixing. The cells were immediately washed with 5ml of standard medium thrice and replaced with a fresh 10ml standard medium (Section II.5.1.1).

II.5.5 Establishment of HeLa cell lines stably expressing a single construct

II.5.5.1 Transfection and Selection Phase

Hela cells were seeded onto a 90mm dish (Merck #402/0322/018) and, when 70% confluency was reached, were transfected with the construct of interest using Effectene reagent (Section II.5.2.2). The medium was replaced with standard medium containing

selected antibiotics two days after transfection (see Section II.3 for details) and the medium was changed every 1-2 days until colonies were observed.

II.5.5.2 Subcloning Phase

6ml of trypsin which has been diluted 1:20 with PBS was used to release the cells from adhering to the plate slowly. 24 individual colonies were selected and picked using a P-200 Gilson pipette. The suspended colonies were dispensed into marked wells in a 24-well plate (GreinerBio-one), each well containing 1.5ml medium with selected antibiotics. To ensure that the colonies picked were expressing the fluorescent fusion protein of interest, both the localisation and general fluorescence intensity was checked under the Zeiss Axiovert S25 using a 10x lens before picking. Apart from very bright colonies (in which the protein of interest may be overexpressed), a range of fluorescent intensities was chosen. After 5-6 days, when the cells reached about 80% confluency, they were trypsinised using 100µl 1x trypsin for 4 minutes and neutralised with 5.5ml of medium with selected antibiotics. 5ml of the cell suspension was transferred to a well in a 6-well plate (GreinerBio-one) while the other 0.5ml was transferred to a well, which contained a 13mm coverslip (Merck, thickness=1.5µm) , in a new 24-well plate.

II.5.5.3 Enrichment Phase

After 2-3 days, the cells, or when cells reached 80% confluency, the cells on the coverslip were fixed as described in Section II.8.2.2. The colonies were ranked according to the percentage of homogeneity and the fluorescent level. The colonies were first ranked with the percentage of homogeneity and the highly homogenous colonies were considered further for their fluorescent level. For those proteins where the endogenous protein levels are low such as RNA polymerase I subunit, the colonies of the low, yet detectable, fluorescent level were then chosen for further enrichment. But usually, 3 different colonies

(of low, medium and high fluorescent level) were chosen and frozen stocks were made by growing the cells seeded on the 6-well plate to a 75cm^2 flask. If the cells colony had reached 100% homogeneity, 1% of the cell suspension derived from the 6-well plate was added to 10ml medium with selected antibiotics in a new 90mm dish for further subcloning as in Section II.5.5.2.

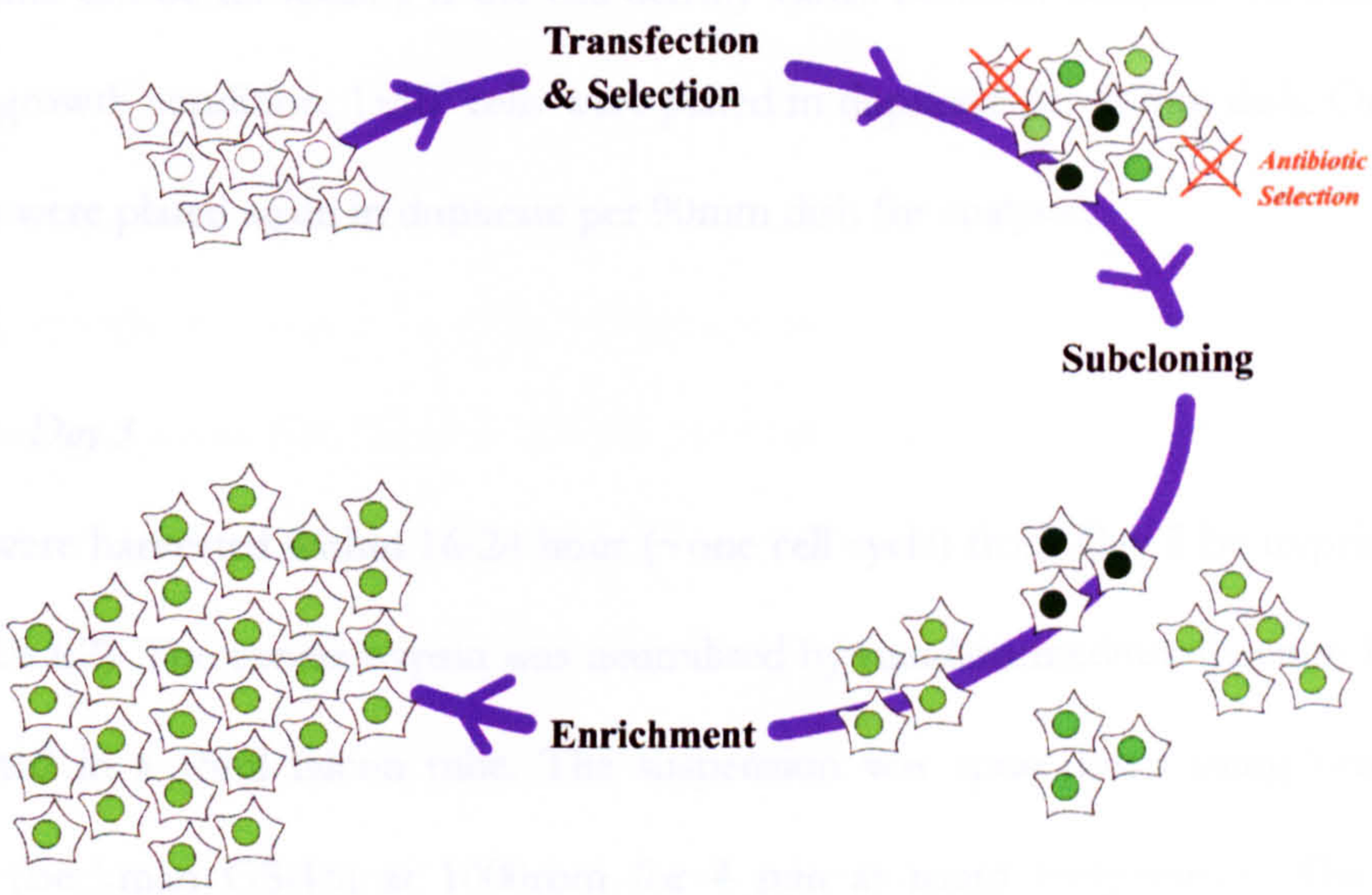


Figure II-3 **Scheme for generation of cell line.** HeLa cell lines were grown to 70% confluency were transfected with chosen plasmid carrying a particular antibiotic-resistance gene. After 48 hours of transfection, the cell lines were selected with antibiotics for 3 weeks. Clones with a range of fluorescent intensities were picked and enriched for future analyses.

II.5.6 Estimation of cell number

The haemocytomer Iused (Mackay & Lynn Ltd # CM175-22; Improved Neubauer version BS748; depth=0.1mm) contains two chambers, each of which contains nine major squares. When filled and coverslipped, each square represents a volume of 0.1mm^3. The cells were trypsinised as in Section II.5.1.2 and two independent samples from the cell suspension was dilution by 1 in 10 and were counted. In each side, only those squares at the corners and the one at the centre were counted to give the number of cells in 1x10^{-3}ml.

The number of cells per ml = no. of cells from 10 squares × dilution factor × 1000.

The haemocytometer and coverslip was cleaned immediately after use by rinsing in MQ H_2O followed by 70% ethanol.

II.5.7 FACS analysis

II.5.7.1 Day 1 & 2

FACS results can be misleading if the cell density varies between samples. To standardise the initial growth condition, $1x10^6$ cells were plated in duplicate per 90mm dish. On day 2, $1x10^6$ cells were plated again in duplicate per 90mm dish for analysis.

II.5.7.2 Day 3

The cells were harvested within 16-24 hour (~one cell cycle) from Day 2 by trypsinisation (Section II.5.1.2). The excess trypsin was neutralised by standard medium (Section II.5.1.1) and collected in a 15ml falcon tube. The suspension was spun down using bench top centrifuge (Beckman GS-15) at 1000rpm for 4 min at room temperature. The excess medium was aspirated and the pellet was dispersed by slight tapping. 1ml of PBS was added for resuspension, followed by another 9 ml of PBS, and the whole cell suspension was spun down again as before. The excess PBS was aspirated and the pellet was again dispersed by slight tapping. For fixation, 1ml of ice-cold 70% Ethanol (in MQ H_2O) was added dropwise while vortexing at a low rate (setting=3). The suspension was left in ice for at least 3 hours until analysis.

II.5.7.3 Just before analysis

The cell suspensions were spun down at 1000rpm for 5 min at 4°C using benchtop centrifuge (Beckman GS-15) and resuspend in 1ml ice-cold PBS. The cell suspensions at this stage were always kept at 4°C to minimise DNA degradation. The cells were counted using a haemocytomer as described in Section II.5.6. After counting the cells, the cells were

once again spun down and resuspend with ice-cold PBS containing 100μg/ml RNase A (Sigma) and 25μg/ml Propidium Iodide (Fluka) such that the final concentration is $5x10^5$ cells/ml. The cell suspension was incubated at 37°C in the dark for 30 minutes. Fluorescence was measured using a FACscan (FL2 channel; 543nm; Becton Dickinson). Cell debris and fixation artifacts were gated out. Data analysis was done using Cell Quest software (Becton Dickinson).

II.5.8 Cryopreservation of cell cultures

Cells were grown at around 70-80% confluency in a $75cm^2$ flask and trypsinised and resuspended as described in Section II.5.1.2. The resultant suspension was spun down at 1000rpm for 4 minutes in room temperature using a benchtop centrifuge (Beckman GS-15). The excess medium was aspirated and the pellet was initially dispersed by slight tapping. 1ml of freezing medium (see below) was added for resuspension using a P-1000 Gilson, followed by an additional 2 ml freezing medium. The cell suspension was aliquoted in 600μl amounts in cryovials (WTS T310-2A; 2ml) and stored in a -80°C freezer for at least 1 week before transferring to liquid nitrogen storage.

Freezing Medium

9ml	Standard medium (Section II.5.1.1)
1ml	DMSO (Sigma #D2650)
10ml	Total

II.5.9 Thawing of cell cultures

The frozen cells retrieved from storage were immediately warmed up at 37°C and resuspended in 5ml of standard medium with appropriate antibiotics in a small flask (GreinerBio-one; base area=$25cm^2$). The cells were allowed to recover in a humidified incubator with 5% CO_2 at 37°C for 4 hours before changing with another 5ml of standard medium. When the cells reached 70-100% confluency, the cells were split as described in Section II.5.1.2.

II.6 Protein analysis

II.6.1 Preparation of nuclear lysate from adherent cells

Cells were seeded in 90mm dishes (Merck #402/0322/18) or 140mm dishes (Merck #402/0322/20) and harvested at 70-80% confluency. The dishes were first rinsed thrice with ice-cold PBS on an ice-cold surface and the excess of PBS were removed by draining to a large beaker. The final remnant of PBS was removed through absorption by a piece of tissue (Kimberly-Clarke #3020030) but great care was made to avoid the contamination of the sample. 0.5ml of Lysis Buffer (as described below) was added per 90mm dish or 1.0ml per 140mm dish and left for 5 minutes on the ice-cold surface. The cells were immediately scraped off the dish with a sterile cell scraper (Greiner #541070) and transferred to 1.5/2ml Eppendorf tubes. To remove clumps of genomic DNA, QIAshredder (QIAGEN #79654) was used according to the manufacturer's suggestion. Essentially, the cells are spun down through a filter to shear the DNA in a 4°C room centrifuge (Eppendorf 5415c; 14000rpm; 2 minute). The filtrate was used for further analysis.

Nuclear Lysis Buffer

1ml	500mM Tris, pH7.5
1ml	5M NaCl
100µl	100% NP-40
0.1g	sodium deoxycholate
100µl	10% SDS
200µl	500mM EDTA (pH8.0)
1	mini-COMPLETE protease inhibitor (Roche #1836170)
10ml	**Total** (with MQ H_2O)

kept at 4°C until use

II.6.2 Preparation of total cell lysates from adherent cells

Cells were seeded in 60mm dishes (Merck #402/0322/12) or 90mm dishes (Merck #402/0322/18) and harvested at 70-80% confluency. After washing thrice with PBS and the removal of excess PBS by aspiration, 200µl 2x LDS (Invitrogen #NP0007) per 60mm dish or 500µl per 90mm dish was added to lyse the cells. The cell lysate was then subjected to QIAshredder as described in Section II.6.1 before use for analysis.

II.6.3 Protein Concentration Measurement

To measure the protein concentration, Coomassie® Protein Assay Reagent Kit (Pierce #23200) was used according to the manufacturer's suggestion with slight modification. Briefly, 10µl of sample was diluted with 90µl MQ H_2O in the 1ml plastic curvette (VMR #307380004) by the addition of 1ml assay reagent. The absorbance at 595nm was then measured in duplicate and subtracted from the reading from the control. To make a proper control, 10µl of buffer in which the sample dissolved was used instead of the sample when measuring the absorbance as described above. The concentration is equal to the absorbance divided by a constant α where α was calculated from a standard curve using multiple defined amounts of BSA.

II.6.4 SDS-Polyacrylamide Gel Electrophoresis (PAGE)

Protein samples were denatured using 1x LDS (Invitrogen #NP0007) and reduced by 0.05M DTT (f/c). The samples were heated at 70°C for 10 minutes before loading onto pre-cast NuPAGE 4-12% Bis-Tris gels (Invitrogen #NP0321-23). Gels were run in either MOPS SDS or MES SDS running buffer (Invitrogen #NP001 and NP002 respectively), depending on which molecular weight range required to be resolved. On the other hand, if a protein of very high molecular weight required resolution, a Tris-acetate gel system would be used instead (Invitrogen #LA0041). The gels were all run at a constant voltage of 200V according to the manufacturer's suggestion.

II.6.5 Western Blotting

II.6.5.1 General

After the electrophoresis, the proteins separated on the gel were transferred to nitrocellulose (Amersham Hybond-C extra) using XCell *SureLock*™ Mini-Cell Blot Module (Invitrogen #EI0001) according to the manufacturer's suggestion on reduced samples. The

transfer was routinely carried out at a constant voltage of 30V for 2 hours using the following transfer buffer.

Transfer buffer
1.45g Tris base
7.2g Glycine
200ml Methanol
1L Total *(with MQ H_2O)*

After the transfer, the blot was routinely stained with a small amount of Ponceau S (Sigma #P7170), which is just sufficient to cover the blot, to reveal the quality of transfer and the overview of the protein pattern. After several washes with MQ H_2O, the blot was recorded using a scanner (Canon; CanoScan D2400U) and stored electronically for future analysis.

II.6.5.2 Stripping of Western Blot

Following immunoblotting (Section II.6.6), the blot could be reused as many as five times by repeated stripping and reprobing. This is instrumental to compare the expression level of endogenous protein and fusion proteins by using sequentially the antibodies raised against the endogenous proteins and the fusion only (such as GFP) on the same blot.

Stripping buffer
0.35ml 14.3M β-mercaptoethanol (BDH)
10ml 10% SDS
12.5ml 0.25M Tris, pH6.8
50ml Total *(with MQ H_2O)*

The blot was incubated with the stripping buffer (above) at 50°C for 30 minutes with occasional shaking and was then washed twice with PBS-T (Section II.1.2). The blot was reblocked with 5% milk in PBS-T overnight before reprobing with antibodies as usual.

II.6.6 Immunological detection of proteins on nitrocellulose filters

After the electrophoresis, the blot was blocked in 5% milk in PBS-T overnight. The blot was probed with the primary antibodies diluted in 5% milk/PBS-T for 1 hour in a sealed plastic bag and then washed thrice with PBS-T every 10 minutes. The primary antibodies

used are listed in Table II-3. The blot was reblocked with 5% milk in PBS-T for at least an hour before adding the secondary antibodies diluted in 5% milk/PBS-T for 30-45 minutes in a sealed plastic bag. The blot was washed thrice with PBS-T every 10 minutes. All the secondary antibodies used in this study are conjugated to horseradish peroxidase (HRP) and could therefore be detected using ECL Plus™ (Amersham Pharmacia biotech #RPN2132) according to the manufacturer's suggestion. The detected blots were either exposed to Kodak film or quantified using Fujifilm Intelligent Dark Box LAS-1000.

II.7 Antibodies

II.7.1 List of antibodies used in this study

Antigen	Name	Properties	IF	IB	Remarks
p80-coilin	5p10	mouse monoclonal	√	√	*lab stock*
SC35		mouse monoclonal	√	√	*purchased from Sigma*
NHPX		rabbit polyclonal	√	√	*generated in this study*
GFP		2 mouse monoclonals	√	√	*purchased from Roche*
Fibrillarin	72B9	mouse monoclonal	√		*lab stock*
Fibrillarin		rabbit polyclonal		√	*gift of Dr. F. Fuller-pace*
Nucleolin	7G2	mouse monoclonal	√	√	*gift of Dr. G. Dreyfuss*
B23		goat polyclonal	√	√	*purchased from Santa Cruz*
UBF		rabbit polyclonal	√		*purchased from Santa Cruz*
B6-2		mouse monoclonal	√		*gift of Dr. P. Cook*
RS6		rabbit polyclonal	√		*gift of Dr. J. Stahl*
U1A	856	rabbit polyclonal	√		*gift of Dr. J. Hamm*
Y12		mouse monoclonal	√		gift of Dr. J. Steitz
p80-coilin	204/10	rabbit polyclonal	√		*lab stock*
nucleoporin	414	mouse monoclonal	√	√	*purchased from Babco*

Table II-3 **List of primary antibodies used in this study.** 'Antigen' indicates the target that the antibodies raised in a particular species as documented under 'Properties'. 'Name' is provided for certain antibodies if a generalised name exists. 'IF' and 'IB' indicate whether the antibodies was used in immunofluorescence and immunoblot, respectively. The sources of the antibodies are indicated under 'Remarks'. Secondary antibodies conjugated with different fluorophores were purchased from Jackson Laboratories.

II.7.2 Generation and characterisation of antibodies

To generate new antibodies raised against peptides, the first thing is to find a peptide region which is likely to be antigenic based on the following criteria: (1) The region is predicted to be exposed to the aqueous phase (according to ProtScale; www.expasy.ch); (2)

the region may contains one or few aromatic residues which increase the antigenicity and (3) the region is unique to the protein of interest. The synthesis of peptides and antisera were carried out commercially by Eurogenetics. Briefly, the peptide region was chemically synthesized and conjugated to a carrier at either the N-terminus or C-terminus before injecting into two rabbits. The pre-immune serum and serum after different round of extraction were received and checked on a blot of total lysate of HeLa cell as described in Section II.6.6. Serial dilution of each serum was compared with the corresponding preimmune serum and the best dilution was selected empirically based on the highest specificity and the lowest background. The following peptides in bold were chosen in this study:

NHPX
MTEADVNPKAYPLADAHLTKKLLDLVQQS**CNYKQLRKGANEATKT**LNRGIS
EFIVMAADAEPLEIILHLPLLCEDKNVPYVFVRSKQALGRACGVSRPVIACSVTIK
EGSQLKQQIQSIQQSIERLLVMTEADVNPKAYPL

II.7.3 Immunoprecipitation Protocol

II.7.3.1 Preparation

In this study, mouse monoclonal antibodies raised against GFP (Roche #1814460) were used and hence protein-G sepharose bead (Pharmacia Biotech #17-0618-01) were chosen for immunoprecipitation due to the higher affinity of mouse antibodies to protein-G compared to protein-A. The sepharose bead preswollen in 20% ethanol were washed with 1ml PBS thrice before making a slurry with 1:1 ratio of beads and PBS. 25µl anti-GFP (0.4mg/ml) was pre-incubated with 50µl protein G-sepharose beads in a 1.5ml Eppendorf tube, shaking at 1000-1200rpm on a electronic vibrator (IKA VIBRATX) overnight at 4°C.

II.7.3.2 Precipitation

On the second day, the lysate was first precleared with PBS-washed protein-G sepharose beads in a 1.5ml Eppendorf tube for 30 minutes at 4°C on a rotating wheel (Labinco

VX2E 528). After the pre-clearing step, the protein concentrations of different samples were measured (Section II.6.3) and normalised. Different pre-cleared lysates of the same protein concentration were then added to the anti-GFP/bead premade the previous day and incubated at 4°C for 4 hours or overnight on a rotating wheel.

II.7.3.3 Washing

The excess lysates were removed by centrifugation at 4°C using Eppendorf 5415c microcentrifuge (14,000 rpm, 1min). The supernatant was kept for further analysis and the pellets were washed with PBS containing protease inhibitor cocktail Complete thrice. The proteins from the pellets were released by adding 1x LDS and boiled at 70°C for 10 minutes and loaded onto a gel for analysis (Section II.6.4).

II.8 Microscopic analysis

II.8.1 Hardware Specification

The following microscopes were used in this study:

Model	Purposes
Axiovert S25	Routine examination of cell culture during propagation and quality control during the isolation of nucleoli from cultured HeLa cells
Axiovert S100	Examination of fluorescence level for both transient and stable transfection and for microinjection purposes (live cell).
Axioplan	Screening of fluorescence level in each clones of stable cell line (fixed cell)
DeltaVision	High resolution of live cell and fixed cell studies
LSM 410	High resolution of live cell and fixed cell studies
LSM 510	High resolution of live cell studies; especially photobleaching experiments

Table II-4 **List of microscope used in this study.**

II.8.1.1 DeltaVision Restoration Microscope Configuration

For obtaining high resolution images, a Zeiss-DeltaVision Restoration microscope (Applied Precision, Inc) equipped with a three dimensional motorised stage was used. Two CCD cameras were used for capturing the fluorescence images according to their readout speed and quality of the data:

	Fixed cell	Live cell
Camera	Photometrics CH350	MicroMax:13000YHS
Read noise	Lower (6e- RMS)	Higher (8e- RMS)
Readout speed	Lower (0.5 MHz)	Higher (5 MHz)
Quantum efficiency	max at 580-650 nm	Max at 400-500 nm

Table II-5 **CCD cameras used in DeltaVision Restoration microscope**

Microscope lamp shutter, focus movements and correct filter combinations are controlled by a Silicon Graphics O_2 workstation using Softworx software. Images were collected using the 100x/NA1.4 Plan-Apochromat objective and recorded using a binning of 2x2 on CH350 CCD camera (fixed cell studies) and 3x3 on Micromax camera (live cell studies) , respectively. The immersion oils (Applied Precision Inc) used had a refractive index of 1.514 for fixed cell and 1.518 for live cells. To distinguish different fluorophores, a combination of filter wheels and stationery beam splitters were used to capture the reflection and transmittance of more than two wavelengths and ensure minimal movement of microscope:

Filter/Beam splitter	Excitation	Emission	Fluorophores
DAPI/PC	P360/W 40	P457/W 50	DAPI
FITC /PC	P490/W 20	P528/W 38	Fluorescein (FITC), GFP, YFP
RD-TR-PE/PC	P555/W 28	P617/W 73	Rhodamine, Texas Red, Cy3
Cy5/PC	P640/W 20	P685/W 40	Cy5
CFP/JP4	P436/W 10	P470/W 30	CFP
YFP/JP4	P500/W 14	P535/W 30	YFP

Table II-6 **Filter set camera used in DeltaVision Restoration microscope.** For excitation and Emission, P indicates the wavelength peak and W indicates the full band width of the filter.

Two stationery beam splitters were used: Polychronic (PC) and JP4 (Chroma technologies #86000 and 86002, respectively) for recording the fluorescence of a single fluorophores whilst preventing bleedthrough from other fluorophores.

II.8.1.2 Zeiss LSM510 Confocal Microscope Configuration

Three or more dimensional imaging was performed on a custom built Zeiss LSM510 (EMBL, Heidelberg) equipped with a z-scanning stage (HRZ 200) for fast 4D acquisition

using a Plan Apochromat 63x DIC oil immersion objective. Triple-colour imaging of CFP, YFP and dsRed/HcRed was achieved by alternating the 413nm Kr, 514nm Ar and 543nm HeNe laser for selective excitation. Bidirectional scanning mode was used to achieve a high scan speed (0.88μs per pixel) without losing much of the quality. Detection for all channels was performed on the same photomultiplier in the microscope, where the emissions were first split with a primary NFT 560 dichroic to a LP560 emission filter (dsRed), and again with a secondary NFT 505 dichroic to a LP 525 (YFP) and a BP 440-505 emission (CFP), respectively (Chroma Technology). Crossover with this setup was below 1% in both channels, with minimized loss of emission from three fluorophores. The configuration required minimal mechanical movement other than the *xy* and *z* scanner. Several macros were written by Mr. G. Rabut (EMBL, Heidelberg) to ensure fast acquisition after photobleaching, multi-point, multi-wavelength time-series scanning and autofocussing during multi-point time-series.

II.8.2 Fixed cell analysis

To capture a snapshot of the cell activity *in situ*, cells are commonly fixed by the following three general methods:

II.8.2.1 Methanol/Acetone fixation

This method is based on dehydration and is useful for exposing buried epitopes. However, it is difficult to preserve the morphology after such a harsh treatment. The cells were fixed by adding 1:1 ice-cold methanol/acetone for 5 minutes, followed by washing twice with 1x PBS.

II.8.2.2 Paraformaldehyde fixation

The cells were first washed thrice with 1x PBS quickly and gently before fixing with 3.7% paraformaldehyde (PFA) in cytoskeletal (CSK) buffer (see below) for 10 minutes. The cells were then washed thrice with 1x PBS. This method is based on the rapid cross-linking reagent PFA to stop the cell activity. Generally, the cell morphology is preserved using this method; however, because of the crosslinking, some of the antibodies cannot access and react to the epitope.

2x PFA/CSK fix

8ml	0.5M PIPES pH6.8
8ml	5M NaCl
47ml	2.55M Sucrose
1.2ml	1M $MgCl_2$
1.6ml	0.5M EDTA
80ml	**5x CSK buffer** *(with MQ H_2O)*
100ml	4x PFA
200ml	**Total** *(with MQ H_2O)*

aliquoted into 5ml per 15ml tubes and kept at -20°C
add 5ml of MQ H_2O before use

To make 4x PFA, 16g of paraformaldehyde (BDH) was weighed in a beaker and 80ml MQ H_2O was added. The mixture was constantly stirred on a heated plate (50-60°C) while the pH was brought up to 7 by adding 10M NaOH dropwise. The pH was checked constantly with pH papers with a range from pH 0-14 (Whatman) and the final volume was made up to 100ml using MQ H_2O.

II.8.2.3 Cell fixation for mitotic studies

However, both methods mentioned above (Sections II.8.2.1-II.8.2.2) are not good at preserving the mitotic cell morphology and the following protocol was used.

37% PFA

1.85g	PFA
3.5ml	MQ H_2O
10µl	10M KOH
5.0ml	**Total**

2x PHEM buffer

18.14g	PIPES
6.5g	HEPES
0.99g	$MgSO_4$
3.8g	EGTA
500ml	**Total** *(with MQ H_2O)*

pH to 7.0 with 10M KOH
filtered and stored at 4°C

Fixation Buffer

25ml	2x PHEM buffer
5ml	37%PFA
50ml	**Total** *(with MQ H_2O)*

The 37% PFA must be made fresh every time before use. To dissolve the materials, they were mixed in a 50 ml Falcon tube with the cap loose in a 60-80°C water bath, with occasionally swirling for no longer than 5 minutes. The fixation buffer was warmed to 37°C and added directly to the cells slowly after pouring the media out of the dish. The cells was fixed for 10 minutes before being washed thrice with 1x PBS very gently (adding the PBS on the wall slowly).

II.8.2.4 Immunofluorescence

Apart from those cells fixed by methanol/acetone (Section II.8.2.1), the cells were first permeabilised before immunofluorescence by the addition of 1% Triton X-100 (BDH) diluted in 1xPBS for 10 minutes with (Section II.8.2.2) or without (Section II.8.2.3) shaking at room temperature. The cells seeded on coverslips were washed with PBS-T once before blocking with 100µl of 1% donkey serum (Jackson Laboratories) for 15 minutes in a humidified chamber. The serum was removed by draining on a piece of tissue (Kimberly-Clarke #3020030) and 100µl primary antibody was added immediately on the coverslip, leaving 1 hour for incubation. This was then washed thrice with PBS-T every 10 minutes before reblocking with 100µl of 1% donkey serum for 15 minutes. Secondary antibodies

was applied for 30 minutes and washed thrice every 5 minutes before mounting (see Section II.8.2.6). All the washing steps in all experiments except for mitotic studies were carried out on a shaker at room temperature (30 rev/min; StuartScientific platform shaker STR6).

II.8.2.5 Cell staining involving 2'-O-methyl RNA

For 2'-*O*-methyl RNA hybridization (Carmo-Fonseca et al., 1992), cells were permeabilised with 0.5% Triton-X-100 in CSK buffer containing protease inhibitor cocktail Complete on ice for 3 minutes and then were fixed in freshly prepared 3.7% paraformaldehyde in CSK buffer for 10 min at room temperature. Cells were washed 3 times in PBS, 1 time in 6X SSPE and prehybridised with 6x SSPE/5x Denhardt's solution containing yeast tRNA (0.5mg/ml) for 15 minutes. Cells were then hybridised with the same buffer with biotinylated 2'-*O*-methyl antisense oligonucleotide probe (2 μM) for 30 minutes and then were washed 3 times in 6x SSPE and rinsed with avidin wash buffer (0.03M HEPES, pH 7.9, 0.15M KCl, 0.05 % Tween-20, 1% Donkey serum) before incubating with TexasRed-conjugated avidin DCS (Vector Labs #A2016) at 2μg/ml for 30 minutes. They were then washed and mounted on slides for microscopic studies (see Section II.8.2.6).

20x SSPE buffer
175.3g NaCl
27.6g $NaH_2PO_4.H_2O$
7.4g EDTA
800ml MQH_2O
pH to 7.4 with NaOH
adjust to 1L and autoclave before use

50x Denhardt's reagent
5g Ficoll
5g Polyvinylpyrrolidone
5g Bovine Serum Albumin (BSA)
500ml MQH_2O

Probes used
U3 probe 1011: 5' – *C*CUUUCGGUGCUC*C*C – 3'
U4 probe 1012: 5' – *C*CUGCCACUGCGCAAAGCU*C*C – 3'
* denote biotinylated sites.

II.8.2.6 Mounting

Before the fixed cells were analysed on the microscope, the coverslips were mounted onto the slide using either Mowiol/Dabco or Vector Shield (Vector Labs #H1000). A glycerol base mounting reagent such as Vector Shield is required for the right refraction properties in using the DeltaVision wide field fluorescence microscope and coverslips were sealed with nail polish.

II.8.3 Live Cell analysis

Both the temperature and the medium contents were instrumental in setting up the live cell analyses.

II.8.3.1 Temperature control devices

The devices were setup according to the manufacturer's suggestions:

Model	Purposes
LabtekII Chamber (Naperville, IL)	Used in LSM510 (EMBL, Heidelberg) for ultrafast 4D confocal imaging and the temperature was kept using a home-built blower to keep the chamber at 37°C.
POC Chamber (Bachofer)	Fitted for LSM410 and LSM510 for live cell studies and the temperature of the chamber was electronically kept at 37°C by the in-built devices.
FCS2 Chamber/ Objective heater (Bioptech)	Fitted for DeltaVision Restoration microscope and the temperature of the chamber and the objective was maintained individually at 37°C for maximal growth, especially for mitotic studies

Table II-7 **Temperature control devices used in this study**

Cells were grown on 42-mm glass coverslip (no. 1; Helmut Sauer) for POC chamber, 40 mm diameter coverslips 1.5 thick for the FCS2 chamber or directly on LabtekII Chamber one day prior to the live cell experiment. Imaging conditions were dependent on the microscopy used (Section II.8.1)

II.8.3.2 Medium

A special medium that does not contain Phenol Red was ordered (Invitrogen #11880-028); otherwise, all medium is set up as in the general growing condition described in Section II.5.1.2). Usually, HEPES was added to give a final concentration of 20mM to buffer any pH change due to the release of CO_2 during analysis. In mitotic studies, 20% FCS was commonly used with the addition of L-ascorbic acid (0.5mg/ml). For analysis that lasts for more than 8 hours in a FCS2 chamber on a DeltaVision microscope, a perfusion chamber was set up such that fresh nutrients were supplied and old medium were replaced at a flow rate of around 0.2-0.5ml/hour via a peristaltic pump, according to the manufacturer's suggestion.

II.9 References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403-10.

Carmo-Fonseca, M., R. Pepperkok, M. Carvalho, and A. Lamond. 1992. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J Cell Biol.* 117:1-14.

Sambrook, Fritsch, and Maniatis. Molecular Cloning - a laboratory manual.

CHAPTER III

THE NUCLEOLAR PROTEOME

"Applying an updated version of a 1960s technique, they treated human nuclei with sugar and sound waves and, using a centrifuge, separated the relatively dense nucleolus from its gelatinous home in the nucleus."

~ News of the week, Science 18th January, 2002 ~

III. The Nucleolar Proteome

III.1 Introduction

The nucleolus is known to be the site of ribosome subunit production. However, recent studies suggest that it may play additional roles. For example, several classes of proteins, including tumour suppressors, splicing and cell-cycle factors and viral proteins will accumulate in the nucleolus, either under certain metabolic conditions, or else at specific stages of the cell cycle (see Chapter I). The biological roles of these facultative nucleolar interactions remain to be uncovered. To understand more about the functions of the human nucleolus, our group, in collaboration with Professor Matthias Mann and co-workers (University of Odense, Denmark), used mass spectrometry to identify the protein components of nucleoli isolated from HeLa cells and part of the work was published (Andersen et al., 2002), which is attached in the Appendix Section.

Nearly 300 nucleolar proteins were identified using MALDI-TOF and nanoelectrospray mass spectrometry on proteins separated by a combination of 1D and 2D gel analyses. Recently, we have extended our coverage of nucleolar proteins to 400 members by performing LC-MS/MS. Approximately 12% of the identified proteins were previously known to be nucleolar in human cells (Section III.3 for details). Surprisingly, ~30% represented either novel or uncharacterised proteins. As expected, nearly all of the known housekeeping proteins required for ribosomal biogenesis were identified in our analyses. So, the immediate question was: what are the remaining 88% proteins doing in the nucleolus? In this chapter, I present a bioinformatics analysis of the nucleolar proteome and suggest a novel way of assigning functions for uncharacterised proteins by mining of transcription profiles, protein-protein interaction data and the currently available, completed genome data from model organisms.

III.2 Purification, Analysis and Verification

Due to their inherent high density, nucleoli could be isolated by sucrose gradient centrifugation following the disruption of nuclei by sonication (Figure III-1). The purity of the sample was examined by a combination of conventional light microscopy, electron microscopy and immunoblot (Andersen et al., 2002). Immunoblot analyses showed that the sample contained nucleolar markers such as fibrillarin and nucleolin but not non-nucleolar NUP62 protein, a component of nuclear pore complex (Andersen et al., 2002). Moreover, the morphology and localisation of subnucleolar domain markers in isolated nucleoli were shown to be indistinguishable from those *in situ* at light and electron microscopy level (Andersen et al., 2002). The proteins in the sample were separated on either a 1D or 2D polyacrylamide gel, followed by in-gel trypsinisation, and the peptides were analysed by tandem mass spectrometry (MS/MS) to identify not only the peptide masses, but also their amino acid compositions (Chapter I). Alternatively, the isolated

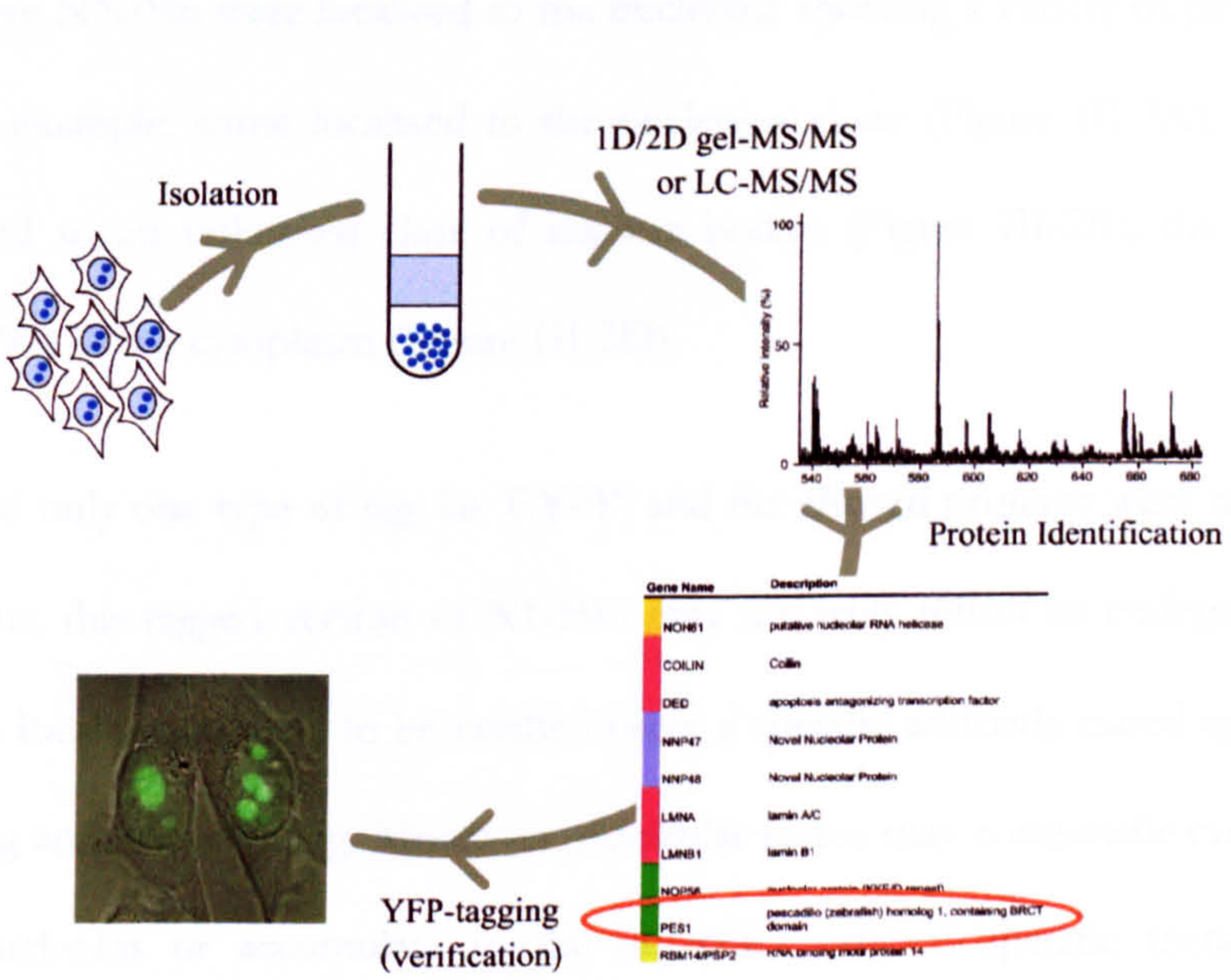


Figure III-1 **Scheme of the characterisation strategy**. Nucleoli were first isolated by sonication and fractionation using a sucrose gradient. The proteins were either separated on a polyacrylamide gel prior to trypsinisation, or first cleaved into peptides before separating using a liquid chromatography. The resultant peptides were analysed by tandem mass spectrometry to identify their masses and amino acid compositions (Chapter I). The information is searched against either the NCBI protein database or human genome database to find the corresponding proteins. To verify the proteins are genuinely nucleolar, they are tagged with YFP and transiently expressed in HeLa cells to examine their localisation.

nucleoli were solubilised, then trypsinised in solution and the resulting peptides were separated through liquid chromatography before analysing by tandem mass spectrometry (LC-MS/MS). The identified peptides from both 1D/2D gel-MS/MS and LC-MS/MS were then used to interrogate the NCBI database and, in some cases, the human genome database to identify the corresponding proteins. The strategy for our nucleolar proteome characterisation is summarised in Figure III-1.

To verify the newly identified components are genuinely nucleolar, I cloned the cDNAs for 10 novel or uncharacterised proteins (namely, NNP62, NNP43/SAZD, NNP42/PWP1, NHPX, PES1, DDX18, NNP51, NNP38, DDX10, NNP46; Table III-3) and inserted into the pEYFPC1 vector (Clontech) to produce a fusion protein of yellow fluorescent protein (YFP) at the amino terminus of each protein. The proteins selected were chosen to include a range of sizes, pI values and motifs. Following transient transfection and expression in HeLa cells, all except NNP46 were localised to the nucleolus showing a variety of patterns (Figure III-2). For example, some localised to the nucleolus alone (Figure III-2A), while others also localised to an unknown class of nuclear bodies (Figure III-2B), the Cajal bodies (Figure III-2C) or the cytoplasm (Figure III-2D).

Because I have used only one type of tag, i.e. EYFP, and the chosen proteins were tagged at only one terminus, this tagged version of NNP46 may not truly reflect its endogenous localisation and the localisation need to be verified using a specific antibody raised against this protein or using an alternative tag. Moreover, nucleolar factor may continually cycle in and out of the nucleolus or accumulate to the nucleolus only at specific metabolic condition or else at specific cell cycle stage (Fox et al., 2002). Therefore, I am confident that the majority, if not all, of the proteins identified, including those encoded by novel genes, are *bona fide* nucleolar factors (Andersen et al., 2002).

To facilitate a comparison of these localisation patterns with known subnucleolar domains (Chapter I), three HeLa cell lines were generated, which label, respectively, the fibrillar centres (EYFP-RPA39), dense fibrillar components (EYFP-FIB) and granular components (EYFP-B23). Surprisingly, a novel protein, NNP43/SAZD, localised to an unidentified subcompartment of nucleoli (Figure III-2E-H). The subcompartment is within the dense fibrillar compartment (Figure III-2F) but does not overlap with the fibrillar centre (Figure III-2G). Interestingly, the pattern is colocalised with SUMO-I within the nucleolus (Figure III-2H); however, the role of this subnucleolar domain remains to be elucidated.

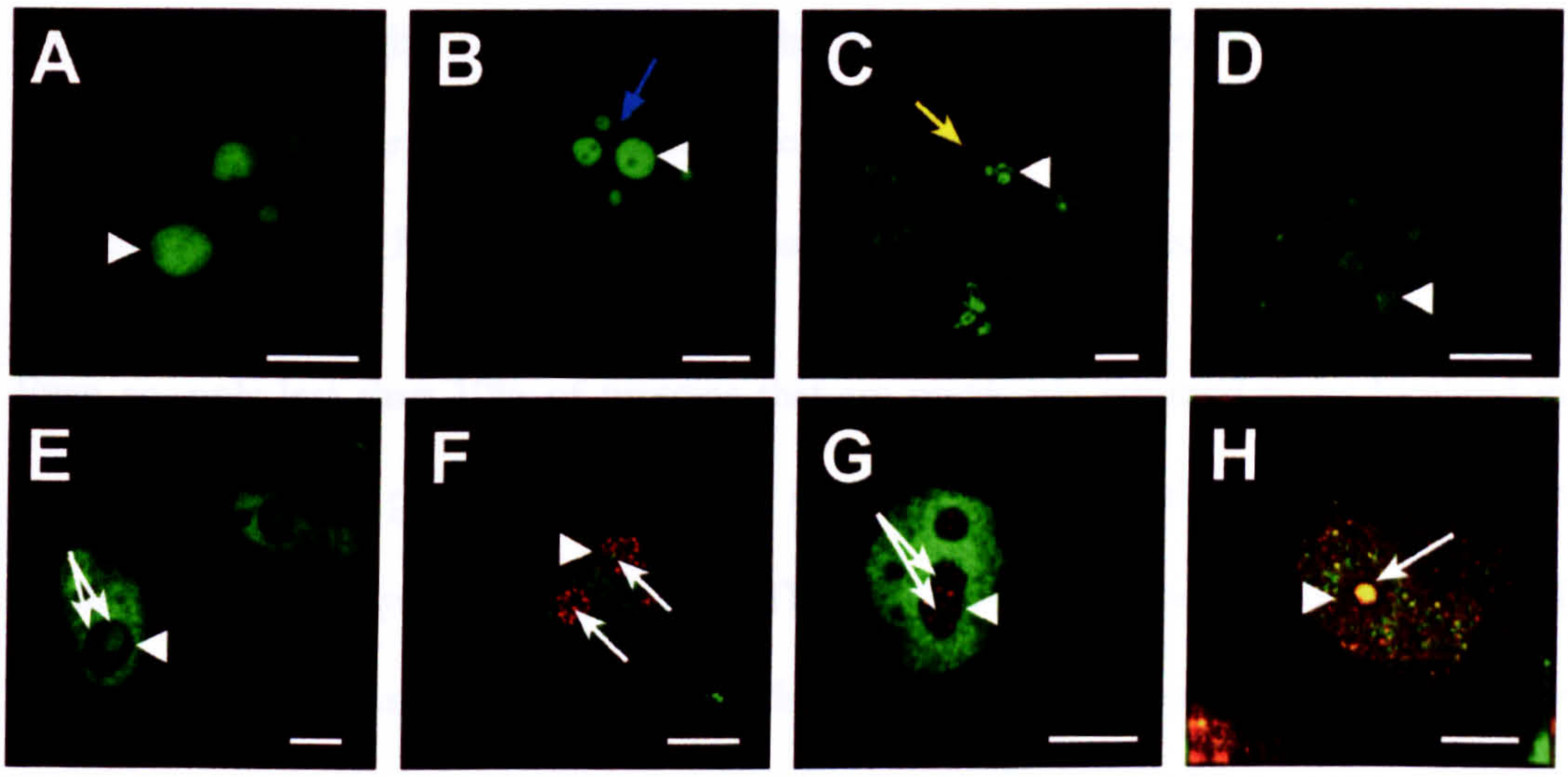


Figure III-2 **Characterisation of novel nucleolar protein localisation.** HeLa cells were fixed after 16 hours of transient transfection using plasmid that expresses (A) EYFP-PES1; (B) EYFP-NNP51; (C) EYFP-NHPX; (D) EYFP-PWP1; (E) EYFP-NNP43/SAZD. To identify the intranucleolar localisation of EYFP-NNP43/SAZD, EYFP-SAZD were transiently transfected to HeLa cells stably transfected with (F) ECFP-FIB and (G) ECFP-RPA39. The transfection condition in (H) is the same as (E) but the transfected cells were further stained with an anti-SUMO1 antibody. White arrowheads indicate nucleoli; blue and yellow arrows indicate unidentified nuclear body and Cajal body respectively; white arrows indicate the intranucleolar localisation of NNP43/SAZD. Scale bar = 5μm.

III.3 Expansion of Known Nucleolar Proteome

In order to compare our nucleolar proteome with the list of previously published nucleolar proteins, it was first necessary to define this comprehensive list. Known nucleolar proteins were identified and extracted from PubMed (1978-2002) using MESH keywords "amino

acid sequences" and word "nucleol*" in the abstract field. 123 nucleolar proteins were thus identified; however, 2 of these proteins were not annotated with amino acid sequences and hence were not compared with our mass spectrometry results. Table III-1 lists the published, known nucleolar proteome as defined by this search.

Examining Table III-1, I note that more than 90% human proteins that were previously reported to be involved in ribosomal biogenesis were identified multiple times, both from the 1D/2D gel and the LC approaches (Table III-1a, d). In general, the LC-MS/MS approach appears more sensitive than the 1D/2D gel approach in detecting peptides and thus gave rise to a higher coverage of nucleolar proteins, indicated by ticks appearing in the LCMS column rather than the CB2002 column (see also Table III-3). ~70% of the proteins identified by 1D/2D gel-MS/MS analyses were also detected by the LC-MS/MS.

Even so, only less than 10% of proteins that are previously reported to be localised in nucleoli during only part of the cell cycle, or under certain metabolic conditions, were identified in our analyses (Table III-1b-c). This may be because the nucleoli we analysed were isolated from unsynchronized HeLa cells and hence facultative nucleolar proteins may only constitute a minor fraction of the proteins isolated. Although LC-MS/MS has already increased the sensitivity of detection, it may still not be sensitive enough to detect very low abundance factors. Future analyses will therefore focus on nucleoli isolated at specific cell cycle stage and under particular metabolic conditions. This should help to increase the total coverage of nucleolar proteins.

Table III-1 **Known Nucleolar Proteome.** Table III-1a shows nucleolar proteins identified in more than one mass spectrometry analyses and most of these proteins, e.g. BRIX, were also found in another recent proteomic analysis of the human nucleolus (Scherl et al., 2002). Proteins known to be localised in nucleoli during only part of the cell cycle and/or under certain metabolic conditions are shown in Table III-1b and Table III-1c respectively. Other known proteins localised in nucleoli are shown in Table III-1d. The column "Accession" shows the corresponding NCBI GI number of the protein. A tick in the column "CB2002" and "LCMS" signifies the identification of the protein in either 1D/2D gel-MS/MS approach (Andersen et al., 2002) or in the LC-MS/MS approach, respectively. The literature regarding the nucleolar localisation of each protein is documented under the column "literature". Note that some proteins fall into multiple categories.

Table III-1a Known nucleolar Proteins found in more than one mass spectrometry analyses

Protein	Name	Accession	CB2002	LCM9	Literature
BLM	Bloom syndrome	4557365	√	√	[1], Table III-1b
BRIX	BRIX	19311012	√		[2]
CGI-94	CGI-94 protein	7705809	√	√	[3]
DDX21	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21	13787209	√	√	[4]
DDX5	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68kD)	4758136	√	√	[5], Table III-1b
DKC1	dyskeratosis congenita 1, dyskerin	4503337	√	√	[6]
FBL	fibrillarin	12056465	√	√	[7]
H2BFE	H2B histone family, member E	4504263	√	√	[8]
hPOP1/KIAA0061	KIAA0061 protein	23097292	√	√	[9]
HSPA1B	heat shock 70kD protein 1B	5123454	√		[10], Table III-1c
HUMAUANTIG	nucleolar GTPase	3334278	√	√	[11]
Ku70/G2ZP1	thyroid autoantigen 70kDa (Ku antigen)	4503641	√	√	[12]
MKI67	antigen identified by monoclonal antibody Ki-67	19923217	√	√	[13]
MKI67IP	MKI67 (FHA domain) interacting nucleolar phosphoprotein	21314753	√	√	[14]
MPHOSPH10	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein)	2230673	√	√	[15]
NCL	nucleolin	4885511	√	√	[16]
NHP2L1	NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)	4826860	√	√	[17]
NOH61	putative nucleolar RNA helicase	9506931	√	√	[18]
NOL1	nucleolar protein 1 (120kD)	5453792	√	√	[19]
NOL5A	nucleolar protein 5A (56kD with KKE/D repeat)	5453794	√	√	[20]
NOLC1	nucleolar and coiled-body phosphoprotein 1	4758860	√	√	[21]
NOP5/NOP58	nucleolar protein NOP5/NOP58	7706254	√	√	[22]
Nop52/D2182056E	DNA segment on chromosome 21 (unique) 2056 expressed sequence	4503247	√	√	[23]
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	10835083	√	√	[24]
Nrap	nucleolar RNA-associated protein	18644728	√	√	[25]
PMSCL1	polymyositis/scleroderma autoantigen 1 (75kD)	4826922	√	√	[26]
PMSCL2	polymyositis/scleroderma autoantigen 2 (100kD)	4505917	√	√	[27]
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isoform	4506007	√	√	[28]
PTBP1	polypyrimidine tract binding protein 1	4506243	√	√	[29]
RPL22	ribosomal protein L22	4506813	√	√	[30]
RPL5	ribosomal protein L5	14591909	√	√	[31]
RPL6	ribosomal protein L6	16753227	√	√	[32]
RPL9	ribosomal protein L9	15431303	√	√	[33]
RPS6	ribosomal protein S6	17156044	√	√	[34]
SURF6	surfeit 6	19557702	√	√	[35]
TCOF1	Treacher Collins-Francaschetti syndrome 1	4507411	√	√	[36]
TOP2B	topoisomerase (DNA) II beta (180kD)	19913408	√	√	[37]
U3-55K	U3 snoRNP-associated 55-kDa protein	4759276	√	√	[38]
UBTF	upstream binding transcription factor, RNA polymerase I	7657671	√	√	[39]

Table III-1b Known proteins localized in nucleoli during part of the cell cycle

Protein	Name	Accession	CB2002	LCM9	Literature
S phase					
BLM	Bloom syndrome	4557365	√	√	[1]
PCNA	proliferating cell nuclear antigen	4505641			[7]
G2/M					
CENPC1	centromere protein C 1	4502779			[40]
STK18	serine/threonine kinase 18	21361433			[41]
Telophase					
DDX5	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68kD)	4758136	√	√	[5]

Table III-1c Known proteins localized in nucleoli under certain metabolic condition

Protein	Name	Accession	CB2002	LCM9	Literature
Apotosis					
DEDD	death effector domain-containing	14670394			[42]
VIL2	villin 2 (ezrin)	31283			[43]
Growth relased					
ANG	angiogenin, ribonuclease, RNase A family, 5	4557313			[44]
AREG	amphiregulin (schwannoma-derived growth factor)	4502199			[45]
FGF2	fibroblast growth factor 2 (basic)	15451898			[46]
FGF3	fibroblast growth factor 3 (murine mammary tumor virus integration site (v-int-2) oncogene homolog)	4885233			[47]
PTHLH	parathyroid hormone-like hormone	4506269			[48]
Heat shock					
HSP105B	heat shock 105kD	2495344			[49]
HSPA1B	heat shock 70kD protein 1B	5123454	√		[10]
HSPB1	heat shock 27kD protein 1	4504517			[50]
PPID	peptidylprolyl isomerase D (cyclophilin D)	4826932			[51]
Interferon induction					
PRKR	protein kinase, interferon-inducible double stranded RNA dependent	4506103			[52]
SP110a	SP110 nuclear body protein isoform a	17986254			[53]
SP110b	SP110 nuclear body protein isoform b	17986256			[53]
SP110c	SP110 nuclear body protein isoform c	17986252			[53]
UV induction					
ING1	inhibitor of growth family, member 1	19923771			[54]
RAD17	RAD17 homolog (S. pombe)	4506383			[17]
Others					
ARL4	ARF-like protein 4	5031603			[55]
CALR	calreticulin	4757900			[56]
CBFA2T3	core-binding factor, runt domain, alpha subunit 2; translocated to, 3	20127532			[57]
DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1	1705473			[58]
GAS41	glioma-amplified sequence-41	5729836			[59]
MDM2	Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse)	4505137			[60]
OXR1	oxidation resistance 1	8922241			[61]
p14ARF	p14 ARF	17736294			[62]
PNMA1	paraneoplastic antigen MA1	14719430			[63]
SE20-4	cutaneous T-cell lymphoma-associated tumor antigen se20-4, differentially expressed nucleolar TGF-beta1 target protein (DENTT)	14861044			[64]
TP53	tumor protein p53 (Li-Fraumeni syndrome)	6400738			[65]
YY1	YY1 transcription factor	4507955			[66]
ZFP37	zinc finger protein 37 homolog (mouse)	4507983			[67]
ZNF259	zinc finger protein 259	4506021			[68]

Table III-1d Other known proteins localized in nucleoli

Protein	Name	Accession	CB2002	LCMS	Literature
Ribosomal proteins					
MRPL3	mitochondrial ribosomal protein L3	6005882			[69]
RPS7	ribosomal protein S7	4506741		√	[70]
RNA processing factors					
POP4	POP4 (processing of precursor , S cerevisiae) homolog	5729986		√	[9]
RNASE3	ribonuclease, RNase A family, 3 (eosinophil cationic protein)	4506551			[71]
RPP14	ribonuclease P (14kD)	5902066			[72]
RPP38	ribonuclease P (38kD)	5454026			[72]
SKIV2L	superkiller viralicidic activity 2-like (S. cerevisiae)	20531967			[73]
Rpp21/FLJ22638	hypothetical protein FLJ22638	15080755			[74]
FTSJ2	FtsJ homolog 2 (E. coli)	7019377			[75]
GEMIN4	gem (nuclear organelle) associated protein 4	7657122			[76]
Translation related factors					
SRP19	signal recognition particle 19kD	4507213			[77]
SRP68	signal recognition particle 68kD	7657617			[77]
SRP72	signal recognition particle 72kD	5902124			[77]
DEAD box proteins					
DDX11	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 (CHL1-like helicase homolog, S. cerevisiae)	13787200			[78]
DDX12	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 (CHL1-like helicase homolog, S. cerevisiae)	1517818			[78]
Nucleotide or Nucleic acid binding proteins					
hCAP-C/SMC4L1	SMC4 structural maintenance of chromosomes 4-like 1 (yeast)	21361252		√	[79]
hCAP-H/KIAA0074	KIAA0074 protein	559715			[79]
SMC2L1	SMC2 structural maintenance of chromosomes 2-like 1 (yeast)	5453591			[79]
ERCC5	excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	4503601			[80]
FMR1	fragile X mental retardation 1	4503765			[81]
FXR1	fragile X mental retardation, autosomal homolog 1	4826736			[82]
FXR2	fragile X mental retardation, autosomal homolog 2	4758410			[82]
WRN	Werner syndrome	5739524			[83]
ANP32B	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	5454088			[84]
JAZ	double-stranded RNA-binding zinc finger protein JAZ	6912440			[85]
KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	5730073		√	[86]
SLC29A2	solute carrier family 29 (nucleoside transporters), member 2	10835019			[87]
STAU	staufen, RNA binding protein (Drosophila)	4759176			[30]
TARBP2	TAR (HIV) RNA binding protein 2	19743840			[88]
TERT	telomerase reverse transcriptase	4507439			[89]
TOP3A	topoisomerase (DNA) III alpha	10835218			[90]
ZNF274	zinc finger protein 274	7706507			[91]
RB1	retinoblastoma 1 (including osteosarcoma)	4508435			[92]

Table III-1d Other known proteins localized in nucleoli (cont'd)

Protein	Name	Accession	CB2002	LCMS	Literature
Others					
CDC2	cell division cycle 2, G1 to S and G2 to M	4502709			[93]
PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform	4505997			[94]
PPP1R10	protein phosphatase 1, regulatory subunit 10	4506009			[95]
ASE-1	CD3-epsilon-associated protein; antisense to ERCC-1	6912246		√	[96]
ASNA1	arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)	4757798		√	[97]
COIL	coilin	4758024	√		[98]
HIC p40	human I-mfa domain-containing protein (HIC p40)	14739509			[99]
HSA8591	nucleolar cysteine-rich protein	7657198			[100]
MCRS1	microspherule protein 1	5453694			[101]
NOL3	nucleolar protein 3 (apoptosis repressor with CARD domain)	4505419			[102]
NOL4	nucleolar protein 4	4505421			[103]
SC65	nucleolar autoantigen (55kD) similar to rat synaptonemal complex protein	3183090			[104]
SENP3	sentrin/SUMO-specific protease 3	11245811		√	[105]
SSB	Sjogren syndrome antigen B (autoantigen La)	10835087			[106]
SSRP1	structure specific recognition protein 1	4507241		√	[107]

References:

1 Yankiwski, V., et al., *Nuclear structure in normal and Bloom syndrome cells.* Proc Natl Acad Sci U S A, 2000. 97(10): p. 5214-9.

2 Kaser, A., et al., *Brix from xenopus laevis and brx1p from yeast define a new family of proteins involved in the biogenesis of large ribosomal subunits.* Biol Chem, 2001. 382(12): p. 1637-47.

3 Heese, K., et al., *Characterizing CGI-94 (comparative gene identification-94) which is down-regulated in the hippocampus of early stage Alzheimer's disease brain.* Eur J Neurosci, 2002. 15(1): p. 79-86.

4 Valdez, B C., et al., *A nucleolar RNA helicase recognized by autoimmune antibodies from a patient with watermelon stomach disease.* Nucleic Acids Res, 1996. 24(7): p. 1220-4.

5 Iggo, R.D., et al., *p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts.* Mol Cell Biol, 1991. 11(3): p. 1326-33.

6 Heiss, N S., et al., *X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions.* Nat Genet, 1998. 19(1): p. 32-8.

7 Busch, H., et al., *Novel nucleolar antigens in autoimmune disease.* J Rheumatol, 1987. 14 Suppl 13: p. 70-7.

8 di Padua Mathieu, D., et al., *Differing accessibility in chromatin of the antigenic sites of regions 1-58 and 63-125 of histone H2B.* J Cell Biol, 1981. 91(1): p. 135-41.

9 Pluk, H., et al., *RNA-protein interactions in the human RNase MRP ribonucleoprotein complex.* Rna, 1999. 5(4): p. 512-24.

10 Milarski, K.L., W.J. Welch, and R.I. Morimoto, *Cell cycle-dependent association of HSP70 with specific cellular proteins.* J Cell Biol, 1989. 108(2): p. 413-23.

11 Racevskis, J., et al., *Cloning of a novel nucleolar guanosine 5'-triphosphate binding protein autoantigen from a breast tumor.* Cell Growth Differ, 1996. 7(2): p. 271-80.

12 Francoeur, A.M., et al., *Identification of Ki (Ku, p70/p80) autoantigens and analysis of anti-Ki autoantibody reactivity.* J Immunol, 1986. 136(5): p. 1648-53.

13 Schonk, D.M., et al., *Assignment of the gene(s) involved in the expression of the proliferation-related Ki-67 antigen to human chromosome 10.* Hum Genet, 1989. 83(3): p. 297-9.

14 Takagi, M., et al., *A novel nucleolar protein, NIFK, interacts with the forkhead associated domain of Ki-67 antigen in mitosis.* J Biol Chem, 2001. 276(27): p. 25386-91.

15 Westendorf, J.M., et al., *M phase phosphoprotein 10 is a human U3 small nucleolar ribonucleoprotein component.* Mol Biol Cell, 1998. 9(2): p. 437-49.
16 Peter, M., et al., *Identification of major nucleolar proteins as candidate mitotic substrates of cdc2 kinase.* Cell, 1990. 60(5): p. 791-801.
17 Chang, M.S., et al., *HRad17 colocalizes with NHP2L1 in the nucleolus and redistributes after UV irradiation.* J Biol Chem, 1999. 274(51): p. 36544-9.
18 Zirwes, R.F., et al., *A novel helicase-type protein in the nucleolus: protein NOH61.* Mol Biol Cell, 2000. 11(4): p. 1153-67.
19 Busch, H., et al., *Nucleolar protein P120 and its targeting for cancer chemotherapy.* Boll Soc Ital Biol Sper, 1991. 67(8): p. 739-50.
20 Gautier, T., et al., *Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis.* Mol Cell Biol, 1997. 17(12): p. 7088-98.
21 Pai, C.Y., et al., *Cell-cycle-dependent alterations of a highly phosphorylated nucleolar protein p130 are associated with nucleologenesis.* J Cell Sci, 1995. 108 (Pt 5): p. 1911-20.
22 Lyman, S.K., L. Gerace, and S.J. Baserga, *Human Nop5/Nop58 is a component common to the box C/D small nucleolar ribonucleoproteins.* Rna, 1999. 5(12): p. 1597-604.
23 Savino, T.M., et al., *The nucleolar antigen Nop52, the human homologue of the yeast ribosomal RNA processing RRP1, is recruited at late stages of nucleologenesis.* J Cell Sci, 1999. 112 (Pt 12): p. 1889-900.
24 Yung, B.Y., A.M. Bor, and Y.H. Yang, *Immunolocalization of phosphoprotein B23 in proliferating and non-proliferating HeLa cells.* Int J Cancer, 1990. 46(2): p. 272-5.
25 Utama, B., et al., *Isolation and characterization of a new nucleolar protein, Nrap, that is conserved from yeast to humans.* Genes Cells, 2002. 7(2): p. 115-32.
26 Gelpi, C., et al., *Identification of protein components reactive with anti-PM/Scl autoantibodies.* Clin Exp Immunol, 1990. 81(1): p. 59-64.
27 Bluthner, M. and F.A. Bautz, *Cloning and characterization of the cDNA coding for a polymyositis-scleroderma overlap syndrome-related nucleolar 100-kD protein.* J Exp Med, 1992. 176(4): p. 973-80.
28 Trinkle-Mulcahy, L., J.E. Sleeman, and A.I. Lamond, *Dynamic targeting of protein phosphatase 1 within the nuclei of living mammalian cells.* J Cell Sci, 2001. 114(Pt 23): p. 4219-28.
29 Huang, S., et al., *The dynamic organization of the perinucleolar compartment in the cell nucleus.* J Cell Biol, 1997. 137(5): p. 965-74.
30 Le, S., R. Sternglanz, and C.W. Greider, *Identification of two RNA-binding proteins associated with human telomerase RNA.* Mol Biol Cell, 2000. 11(3): p. 999-1010.
31 Michael, W.M. and G. Dreyfuss, *Distinct domains in ribosomal protein L5 mediate 5 S rRNA binding and nucleolar localization.* J Biol Chem, 1996. 271(19): p. 11571-4.
32 Thomson, S.R. and S.E. Johnson, *Isolation and characterization of chicken TaxREB107, a putative DNA binding protein abundantly expressed in muscle.* Gene, 2001. 278(1-2): p. 81-8.
33 Vater, C.A., et al., *Ricin A chain can be chemically cross-linked to the mammalian ribosomal proteins L9 and L10e.* J Biol Chem, 1995. 270(21): p. 12933-40.
34 Franco, R. and M.G. Rosenfeld, *Hormonally inducible phosphorylation of a nuclear pool of ribosomal protein S6.* J Biol Chem, 1990. 265(8): p. 4321-5.
35 Magoulas, C. and M. Fried, *Isolation and genomic analysis of the human surf-6 gene: a member of the Surfeit locus.* Gene, 2000. 243(1-2): p. 115-23.
36 Winokur, S.T. and R. Shiang, *The Treacher Collins syndrome (TCOF1) gene product, treacle, is targeted to the nucleolus by signals in its C-terminus.* Hum Mol Genet, 1998. 7(12): p. 1947-52.
37 Zini, N., et al., *Discrete localization of different DNA topoisomerases in HeLa and K562 cell nuclei and subnuclear fractions.* Exp Cell Res, 1994. 210(2): p. 336-48.
38 Pluk, H., et al., *cDNA cloning and characterization of the human U3 small nucleolar ribonucleoprotein complex-associated 55-kilodalton protein.* Mol Cell Biol, 1998. 18(1): p. 488-98.
39 Jantzen, H.M., et al., *Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins.* Nature, 1990. 344(6269): p. 830-6.
40 Pluta, A.F. and W.C. Earnshaw, *Specific interaction between human kinetochore protein CENP-C and a nucleolar transcriptional regulator.* J Biol Chem, 1996. 271(31): p. 18767-74.
41 Hudson, J.W., et al., *Late mitotic failure in mice lacking Sak, a polo-like kinase.* Curr Biol, 2001. 11(6): p. 441-6.
42 Stegh, A.H., et al., *DEDD, a novel death effector domain-containing protein, targeted to the nucleolus.* Embo J, 1998. 17(20): p. 5974-86.
43 Kaul, S.C., et al., *Identification of a 55-kDa ezrin-related protein that induces cytoskeletal changes and localizes to the nucleolus.* Exp Cell Res, 1999. 250(1): p. 51-61.
44 Moroianu, J. and J.F. Riordan, *Identification of the nucleolar targeting signal of human angiogenin.* Biochem Biophys Res Commun, 1994. 203(3): p. 1765-72.
45 Modrell, B., V.L. McDonald, and M. Shoyab, *The interaction of amphiregulin with nuclei and putative nuclear localization sequence binding proteins.* Growth Factors, 1992. 7(4): p. 305-14.
46 Gualandris, A., et al., *Modulation of cell growth and transformation by doxycycline-regulated FGF-2 expression in NIH-3T3 cells.* J Cell Physiol, 1999. 181(2): p. 273-84.
47 Kiefer, P. and C. Dickson, *Nucleolar association of fibroblast growth factor 3 via specific sequence motifs has inhibitory effects on cell growth.* Mol Cell Biol, 1995. 15(8): p. 4364-74.
48 Henderson, J.E., et al., *Nucleolar localization of parathyroid hormone-related peptide enhances survival of chondrocytes under conditions that promote apoptotic cell death.* Mol Cell Biol, 1995. 15(8): p. 4064-75.
49 Parsell, D.A., et al., *Hsp104 is a highly conserved protein with two essential nucleotide-binding sites.* Nature, 1991. 353(6341): p. 270-3.
50 Biggiogera, M., et al., *Localization of heat shock proteins in mouse male germ cells: an immunoelectron microscopical study.* Exp Cell Res, 1996. 229(1): p. 77-85.
51 Owens-Grillo, J.K., et al., *A model of protein targeting mediated by immunophilins and other proteins that bind to hsp90 via tetratricopeptide repeat domains.* J Biol Chem, 1996. 271(23): p. 13468-75.
52 Jeffrey, I.W., et al., *Nuclear localization of the interferon-inducible protein kinase PKR in human cells and transfected mouse cells.* Exp Cell Res, 1995. 218(1): p. 17-27.
53 Welsh, G.I., et al., *Colocalization within the nucleolus of two highly related IFN-induced human nuclear phosphoproteins with nucleolin.* Exp Cell Res, 1999. 250(1): p. 62-74.
54 Scott, M., et al., *UV induces nucleolar translocation of ING1 through two distinct nucleolar targeting sequences.* Nucleic Acids Res, 2001. 29(10): p. 2052-8.
55 Lin, C.Y., et al., *ARL4, an ARF-like protein that is developmentally regulated and localized to nuclei and nucleoli.* J Biol Chem, 2000. 275(48): p. 37815-23.
56 Opas, M., et al., *Regulation of expression and intracellular distribution of calreticulin, a major calcium binding protein of nonmuscle cells.* J Cell Physiol, 1991. 149(1): p. 160-71.
57 Hoogeveen, A.T., et al., *The transcriptional corepressor MTG16a contains a novel nucleolar targeting sequence deranged in t(16; 21)-positive myeloid malignancies.* Oncogene, 2002. 21(43): p. 6703-12.
58 Hattori, H., et al., *Intracellular localization and partial amino acid sequence of a stress-inducible 40-kDa protein in HeLa cells.* Cell Struct Funct, 1992. 17(1): p. 77-86.
59 Mannan, A., et al., *Expression, cellular distribution and protein binding of the glioma amplified sequence (GAS41), a highly conserved putative transcription factor.* Oncogene, 2001. 20(35): p. 4853-63.

60 Rizos, H., et al., *Two arginine rich domains in the p14ARF tumour suppressor mediate nucleolar localization.* Oncogene, 2000 19(26): p. 2978-85.
61 Fischer, H., et al., *C7, a novel nucleolar protein, is the mouse homologue of the Drosophila late puff product L82 and an isoform of human OXR1.* Biochem Biophys Res Commun, 2001.281(3): p. 795-803.
62 Lindstrom, M.S., et al., *Immunolocalization of human p14(ARF) to the granular component of the interphase nucleolus.* Exp Cell Res, 2000 256(2): p. 400-10.
63 Dalmau, J., et al., *Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders.* Brain, 1999. 122 (Pt 1): p. 27-39.
64 Chai, Z., et al., *SET-related cell division autoantigen-1 (CDA1) arrests cell growth.* J Biol Chem, 2001. 276(36): p. 33665-74.
65 Bensinghoff, J., et al., *Two different forms of p53 localized differently within cells of urogenital tumours.* Cancer Lett, 1999. 144(1): p. 55-64.
66 McNeil, S., et al., *Targeting of the YY1 transcription factor to the nucleolus and the nuclear matrix in situ: the C-terminus is a principal determinant for nuclear trafficking.* J Cell Biochem, 1998. 68(4): p. 500-10.
67 Payen, E., et al., *The centromeric/nucleolar chromatin protein ZFP-37 may function to specify neuronal nuclear domains.* J Biol Chem, 1998. 273(15): p. 9099-109.
68 Galcheva-Gargova, Z., et al., *The cytoplasmic zinc finger protein ZPR1 accumulates in the nucleolus of proliferating cells.* Mol Biol Cell, 1998. 9(10): p. 2963-71.
69 Ou, J.H., et al., *Cloning and characterization of a human ribosomal protein gene with enhanced expression in fetal and neoplastic cells.* Nucleic Acids Res, 1987 15(21): p. 8919-34.
70 Annilo, T., et al., *Nuclear import and nucleolar accumulation of the human ribosomal protein S7 depends on both a minimal nuclear localization sequence and an adjacent basic region.* Biochem Biophys Res Commun, 1998.249(3): p. 759-66.
71 Wu, H., et al., *Human RNase III is a 160-kDa protein involved in preribosomal RNA processing.* J Biol Chem, 2000. 275(47): p. 36957-65.
72 Jarrous, N., et al., *Localization in the nucleolus and coiled bodies of protein subunits of the ribonucleoprotein ribonuclease P.* J Cell Biol, 1999. 146(3): p. 559-72.
73 Lee, S.G., et al., *Identification and characterization of a human cDNA homologous to yeast SKI2.* Genomics, 1995. 25(3): p. 660-6.
74 Jarrous, N., et al., *Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P.* Rna, 2001. 7(8): p. 1153-64.
75 Ching, Y.P., et al., *Identification and characterization of FTSJ2, a novel human nucleolar protein homologous to bacterial ribosomal RNA methyltransferase.* Genomics, 2002. 79(1): p. 2-6.
76 Charroux, B., et al., *Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli.* J Cell Biol, 2000. 148(6): p. 1177-86.
77 Jacobson, M.R. and T. Pederson, *Localization of signal recognition particle RNA in the nucleolus of mammalian cells.* Proc Natl Acad Sci U S A, 1998. 95(14): p. 7981-6.
78 Amann, J., V.J. Kidd, and J.M. Lahti, *Characterization of putative human homologues of the yeast chromosome transmission fidelity gene, CHL1.* J Biol Chem, 1997. 272(6): p. 3823-32.
79 Cabello, O.A., et al., *Cell cycle-dependent expression and nucleolar localization of hCAP-H.* Mol Biol Cell, 2001.12(11): p. 3527-37.
80 Kaufmann, J.A., et al., *Multiple nuclear localization signals in XPG nuclease.* Mutat Res, 1996. 363(1): p. 67-75.
81 Willemsen, R., et al., *Association of FMRP with ribosomal precursor particles in the nucleolus.* Biochem Biophys Res Commun, 1996. 225(1): p. 27-33.
82 Tamanini, F., et al., *The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins.* Hum Mol Genet, 2000 9(10): p. 1487-93.
83 Marciniak, R.A., et al., *Nucleolar localization of the Werner syndrome protein in human cells.* Proc Natl Acad Sci U S A, 1998. 95(12): p. 6887-92.
84 Zhu, L., et al., *Cloning and characterization of a new silver-stainable protein SSP29, a member of the LRR family.* Biochem Mol Biol Int, 1997. 42(5): p. 927-35.
85 Yang, M., W.S. May, and T. Ito, *JAZ requires the double-stranded RNA-binding zinc finger motifs for nuclear localization.* J Biol Chem, 1999. 274(39). p. 27399-406.
86 Venables, J.P., et al., *T-STAR/ETOILE. a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis.* Hum Mol Genet, 1999. 8(6): p. 959-69.
87 Williams, J.B. and A.A. Lanahan, *A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein.* Biochem Biophys Res Commun, 1995 213(1): p. 325-33.
88 Eckmann, C.R. and M.F. Jantsch, *Xlrbpa, a double-stranded RNA-binding protein associated with ribosomes and heterogeneous nuclear RNPs.* J Cell Biol, 1997. 138(2): p. 239-53.
89 Yang, Y., et al., *Nucleolar localization of hTERT protein is associated with telomerase function.* Exp Cell Res, 2002. 277(2): p. 201-9.
90 Lin, C.W., et al., *Differential expression of human topoisomerase IIIalpha during the cell cycle progression in HL-60 leukemia cells and human peripheral blood lymphocytes.* Exp Cell Res, 2000. 256(1): p. 225-36.
91 Huang, Z., et al., *Expression of the transcriptional repressor protein Kid-1 leads to the disintegration of the nucleolus.* J Biol Chem, 1999 274(12): p. 7640-8.
92 Bukovsky, A., et al., *Quantitative evaluation of the cell cycle-related retinoblastoma protein and localization of Thy-1 differentiation protein and macrophages during follicular development and atresia, and in human corpora lutea.* Biol Reprod, 1995. 52(4): p. 776-92.
93 Ino, H., et al., *p34cdc2 homologue is located in nucleoli of the nervous and endocrine systems.* Brain Res, 1993. 614(1-2): p. 131-6.
94 Kotani, H., et al., *The delta isoform of protein phosphatase type 1 is localized in nucleolus and dephosphorylates nucleolar phosphoproteins.* Biochem Biophys Res Commun, 1998 249(1): p. 292-6.
95 Krevvi, J.P., et al., *Purification and characterization of p99, a nuclear modulator of protein phosphatase 1 activity.* FEBS Lett, 1997. 420(1): p. 57-62.
96 Whitehead, C.M., et al., *ASE-1. a novel protein of the fibrillar centres of the nucleolus and nucleolus organizer region of mitotic chromosomes.* Chromosoma, 1997. 106(8): p. 493-502.
97 Kurdi-Haidar, B., et al., *Dual cytoplasmic and nuclear distribution of the novel arsenite-stimulated human ATPase (hASNA-I).* J Cell Biochem, 1998. 71(1): p. 1-10.
98 Bohmann, K., J.A. Ferreira, and A.I. Lamond, *Mutational analysis of p80 coilin indicates a functional interaction between coiled bodies and the nucleolus.* J Cell Biol, 1995. 131(4): p. 817-31.
99 Thebault, S., et al., *Sequence requirement for the nucleolar localization of human L-myc domain-containing protein (HIC p40).* Eur J Cell Biol, 2000. 79(11): p. 834-8.
100 Bolivar, J., et al., *Molecular cloning of a zinc finger autoantigen transiently associated with interphase nucleolus and mitotic centromeres and midbodies. Orthologous proteins with nine CXXC motifs highly conserved from nematodes to humans.* J Biol Chem, 1999 274(51): p. 36456-64.
101 Rea, Y., et al., *The 58-kDa microspherule protein (MSP58), a nucleolar protein, interacts with nucleolar protein p120.* Eur J Biochem, 1998. 253(3): p. 734-42.
102 Stoss, O., et al., *Alternative splicing determines the intracellular localization of the novel nuclear protein Nop30 and its interaction with the splicing factor SRp30c.* J Biol Chem, 1999. 274(16): p. 10951-62.
103 Ueki, N., et al., *NOLP identification of a novel human nucleolar protein and determination of sequence requirements for its nucleolar localization.* Biochem Biophys Res Commun, 1998.252(1): p. 97-102.
104 Ochs, R.L., et al., *cDNA cloning and characterization of a novel nucleolar protein.* Mol Biol Cell, 1996. 7(7): p. 1015-24.
105 Nishida, T., H. Tanaka, and H. Yasuda, *A novel mammalian Smt3-specific isopeptidase 1 (SMT3IP1) localized in the nucleolus at interphase.* Eur J Biochem, 2000. 267(21): p. 6423-7.
106 Carmo-Fonseca, M., et al., *Identification of La ribonucleoproteins as a component of interchromatin granules.* Exp Cell Res, 1989 185(1): p. 73-85.
107 Hsu, T., et al., *A Drosophila single-strand DNA/RNA-binding factor contains a high-mobility-group box and is enriched in the nucleolus.* Proc Natl Acad Sci U S A, 1993. 90(14): p. 6488-92.

III.4 Bioinformatics Analyses

III.4.1 General

III.4.1.1 Background

The large amount of data acquired from the proteomic studies requires a systematic way to analyse and integrate it with the information already deposited in the databases publicly available (Table III-2). To facilitate these analyses, several databases were downloaded to our UNIX server (bioinformatics.msiwtb.dundee.ac.uk) and were interrogated with software such as standalone BLAST (Altschul et al., 1990; Altschul et al., 1997) and customized PERL scripts I have written throughout my PhD studies (the scripts are deposited in the accompanying CD). The resulting information was presented in a searchable database written in Macromedia FLASH action script and published together with the Current Biology paper in January 2002 at http://www.dundee.ac.uk/lifesciences/lamonddatabase/ (a simplified version can be found in the accompanying CD).

Website	URL
Retrieval of sequences and databases	
Swissprot	http://www.expasy.ch/
EBI	http://www.ebi.ac.uk/
Entrez	http://www.ncbi.nlm.nih.gov/Entrez/
mRNA expression and genome data	
Unigene	http://www.ncbi.nlm.nih.gov/Unigene/
SAGE	http://www.ncbi.nlm.nih.gov/SAGE/
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/
Literature	
OMIM	http://www.ncbi.nlm.nih.gov/OMIM/
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/
Homologues search	
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
SGD	http://genome-www.stanford.edu/Saccharomyces/
MIPS	http://mips.gsf.de/proj/yeast/CYGD/db/index.html

Table III-2 **Public databases used in this study**

The construction of nucleolar proteome database is summarised as follows: The peptide sequences obtained from the mass spectrometry were assigned to a particular protein sequence either by BLAST or by more commonly, MASCOT (http://www.expasy.ch/). All the sequences were then retrieved by Batch Entrez from the NCBI database. The protein sequences were then assigned to their corresponding gene by BLAST analysis against the Unigene database. The Unigene entry provides a starting point leading to various other relevant databases, such as LocusLink and OMIM, to provide the cognate genomic and literature data, while the information deposited in the Swissprot database provides physical properties of each protein, such as isoelectric point (pI) and Molecular weight (Mw). A simplified version of the database is summarised in Table III-3 and the full table can be found in the accompanying CD.

III.4.1.2 Classification by known motifs and homologues

Due to the large number of novel proteins, accounting for ~30% of the total nucleolar proteome, existing information available from the database annotation is not enough to predict the functions of most of the proteins and hence their motifs and homologues were searched with the aim of getting more information for classification. The search of motifs and homologues was conducted via BLAST against Conserved Domain Database or CDD (Marchler-Bauer et al., 2002) and specific proteome databases of the model organisms including *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Caenorhabditis elegans*. The most abundant motifs were the RNA-binding RRM domain, DEAD box helicase domain and the WD domain. The motif distribution within the human nucleolar proteome is deposited in Supplementary Table III-S1 in the accompanying CD.

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
PRKDC	protein kinase, DNA-activated, catalytic polypeptide	465433	155637	5591		√	TOR1				
MKI67	antigen identified by monoclonal antibody Ki-67	358746	80976	4288		√	MUC1				
DSP	desmoplakin (DPI, DPII)	331776	349499	1832		√	USO1				
NNP73	Novel Nucleolar Protein 73	318426	178614	27340		√	UTP20				
ATR	ataxia telangiectasia and Rad3 related	301367	77613	545			MEC1				
PRP8	PRP8 pre-mRNA processing factor 8 homolog (yeast)	273600	181368	10594		√	PRP8				
GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	264342	75354	10985		√	GCN1				
NNP72	Novel Nucleolar Protein 72	252319	18759	9875		√					
NNP71	Novel Nucleolar Protein 71	242372	285861	55127		√	UTP10				
NUMA1	nuclear mitotic apparatus protein 1	238274	301512	4926		√	USO1				
MDN1	MDN1, midasin homolog (yeast)	231844	76730	23195		*	MDN1				
FER1L3	fer-1-like 3, myoferlin (C. elegans)	229706	234680	26509		√					
CHD4	chromodomain helicase DNA binding protein 4	217991	74441	1108			CHD1				
RRP5	likely ortholog of mouse programmed cell death protein 11	210045	239499	22984	2p	√	RRP5				
U5-200K	KIAA0788 protein	194479	246112	23020		√	BRR2				
RPA190	likely ortholog of mouse RNA polymerase 1-4 (194 kDa subunit)	194191	59475	25885	1	√	RPA190				
TOP2B	topoisomerase (DNA) II beta 180kDa	182663	75248	7155		√	TOP2				
TOP2A	topoisomerase (DNA) II alpha 170kDa	174385	156346	7153		√	TOP2				
NNP70	Novel Nucleolar Protein 70	170570	57730	9816		√					
BLM	Bloom syndrome	159000	36820	641	1,2*	√	SGS1				
MYBBP1A	MYB binding protein (P160) 1a	148850	22824	10514	1	√	POL5				
SMC4L1	SMC4 structural maintenance of chromosomes 4-like 1 (yeast)	147168	50758	10051	1	*	SMC4				
CIT	citron (rho-interacting, serine/threonine kinase 21)	146506	15767	11113		*	SPC110				
NNP69	Novel Nucleolar Protein 69	145807	10848	9790	2	√	BMS1				
TCOF1	Treacher Collins-Franceschetti syndrome 1	144312	301266	6949	1	√	MUC1				
SMC1L1	SMC1 structural maintenance of chromosomes 1-like 1 (yeast)	142301	211602	8243		√	SMC1				
CSPG6	chondroitin sulfate proteoglycan 6 (bamacan)	141542	24485	9126		√	SMC3				
DDX9	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 9 (RNA helicase A, nuclear DNA helicase II; leukophysin)	140877	74578	1660		√	YLR419W				
RAD50	RAD50 homolog (S. cerevisiae)	138432	41587	10111			RAD50				
ADAR	adenosine deaminase, RNA-specific	135967	7957	103		*	TAD1				
NNP68	Novel Nucleolar Protein 68	134598	60103	23223		√	RRP12				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
RFC1	replication factor C (activator 1) 1 (145kD)	128254	166563	5981		*	RFC1				
NNP27	Novel Nucleolar Protein 27	127593	183253	65083			UTP22				
NNP102	Novel Nucleolar Protein 102	122831	11387	57602		*	UBP8				
NNP119	Novel Nucleolar Protein 119	122138	86337	84172		*	RPA135				
NNP97	Novel Nucleolar Protein 97	120401	NA			*	MDN1				
FACTP140	chromatin-specific transcription elongation factor large subunit	119914	14963	11198		*	SPT16				
NNP67	Novel Nucleolar Protein 67	118256	278608	23517	2	√	MTR4				
CBF2	CCAAT-box-binding transcription factor	114072	184760	10153		√	MAK21				
NNP66	Novel Nucleolar Protein 66	113913	71472	55226		√	KRE33				
NNP95	Novel Nucleolar Protein 95	113360	323346	22894		*	DIS3				
ADPRT	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)	113084	177766	142		√					
DDX37	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 37	110202	107382	57647		*	ECM16				
NNP65	Novel Nucleolar Protein 65	109145	274149	27043		√	VRP1				
XRN2	5'-3' exoribonuclease 2	108582	268555	22803		*	RAT1				
NNP64	Novel Nucleolar Protein 64	107366	7482	9904		√	MRD1				
NNP63	Novel Nucleolar Protein 63	106099	33085	10885		√	DIP2				
NNP104	Novel Nucleolar Protein 104	105693	26369	54881		*	YHR085W				
NNP62	Novel Nucleolar Protein 62	102451	79380	5822	*	√	PWP2				
NNP61	Novel Nucleolar Protein 61	101433	170114	10940		√	POP1				
MCM2	MCM2 minichromosome maintenance deficient 2, mitotin (S. cerevisiae)	100926	57101	4171		*	MCM2				
PMSCL2	polymyositis/scleroderma autoantigen 2, 100kDa	100831	75584	5394	1,2	√	RRP6				
DDX10	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 10 (RNA helicase)	100816	41706	1662	2p,*	√	HCA4				
NNP49	Novel Nucleolar Protein 49	99366	175596	134430			UTP21				
NNP87	Novel Nucleolar Protein 87	99269	117487	8602		*	NOP14				
NNP124	Novel Nucleolar Protein 124	98796	88820	51575		*	YDR365C				
NNP76	Novel Nucleolar Protein 76	98666	70582	79039		√	DBP10				
DNM2	dynamin 2	97592	167013	1785		*	DNM1				
NNP111	Novel Nucleolar Protein 111	96801	48712	55035		*	SRP40				
DDX24	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 24	96332	155986	57062		*	MAK5				
EEF2	eukaryotic translation elongation factor 2	95338	75309	1938		*	EFT1				
NNP86	Novel Nucleolar Protein 86	95051	58927	4931		*	YLL034C				
MATR3	matrin 3	94623	78825	9782		*	MDN1				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NOL1	nucleolar protein 1, 120kDa	94078	15243	4839	1,2	√	NOP2				
MCM6	MCM6 minichromosome maintenance deficient 6 (MIS5 homolog, S. pombe) (S. cerevisiae)	92889	155462	4175		*	MCM6				
ABCF1	ATP-binding cassette, sub-family F (GCN20), member 1	91680	9573	23		*	KRE30				
MCM3	MCM3 minichromosome maintenance deficient 3 (S. cerevisiae)	90941	179565	4172		*	MCM3				
DDX15	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 15	90933	5683	1665		√	PRP43				
TOP1	topoisomerase (DNA) I	90726	317	7150		√	TOP1				
HNRPU	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	90479	103804	3192		*	NPL3				
UBTF	upstream binding transcription factor, RNA polymerase I	89406	89781	7343		√	YOR054C				
NNP43/SAZD	Novel Nucleolar Protein 43	89035	114416	10607	*		UTP13				
NNP122	Novel Nucleolar Protein 122	87978	271926	10813		*	UTP14				
NNP109	Novel Nucleolar Protein 109	87864	182669	84128		*	PAC1				
NNP60	Novel Nucleolar Protein 60	87055	194754	27042		√	YIL091C				
NNP48	Novel Nucleolar Protein 48	86605	65234	55661	2p	√	DRS1				
NNP59	Novel Nucleolar Protein 59	85766	274263	55131	2p	√	NOP4				
NNP112	Novel Nucleolar Protein 112	85170	219614	22992		*					
NNP58	Novel Nucleolar Protein 58	84878	134200	26155	2	√	NOC2				
NNP103	Novel Nucleolar Protein 103	84469	85769	30836		*	YLR051C				
NNP57/NGB	Novel Nucleolar Protein 57	83655	75528	29889		√	NOG2				
HSPCB	heat shock 90kD protein 1, beta	83264	74335	3326		*	HSP82				
NNP118	Novel Nucleolar Protein 118	83065	321618	81887		*	LAS1				
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining; Ku autoantigen, 80kD)	82705	84981	7520		*					
GU2	nucleolar protein GU2	82565	7392	79009		√	DBP3				
NNP56	Novel Nucleolar Protein 56	82160	152629	23076		√	RRP1				
KNSL6	kinesin-like 6 (mitotic centromere-associated kinesin)	81313	69360	11004		*	KIP3				
NNP55	Novel Nucleolar Protein 55	81124	26761	25926		√					
SSRP1	structure specific recognition protein 1	81075	79162	6749	1	*	POB3				
MCM4	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)	79718	154443	4173		*	CDC54				
DDX21	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21	79657	169531	9188	1	√	DBP3				
DDX33	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 33	78741	250456	56919		*	PRP22				
MPHOPSH10	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein)	77682	201676	10199		√	MPP10				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NRF	transcription factor NRF	77673	119018	55922		*					
HNRPM	heterogeneous nuclear ribonucleoprotein M	77470	79024	4670		√	GBP2				
NNP54/BOP1	Novel Nucleolar Protein 54	76949	30736	23246	2p	√	ERB1				
NCL	nucleolin	76344	79110	4691	1	√	NSR1				
SFPQ/PSF	splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)	76149	180610	6421		√	LAS17				
ILF3	interleukin enhancer binding factor 3, 90kD	76115	256583	3609		*					
NNP53	Novel Nucleolar Protein 53	75721	90315	23160		√	UTP5				
DDX31	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 31	75433	69331	64794		*	DBP7				
DDX18	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated)	75411	100555	8886	2*	√	HAS1				
NNP52	Novel Nucleolar Protein 52	74503	235498	79954		√	YGR145W				
LMNA	lamin A/C	74139	77886	4000			USO1				
NNP51	Novel Nucleolar Protein 51	73965	215766	23560	2*	√	NOG1				
NOLC1	nucleolar and coiled-body phosphoprotein 1	73720	75337	9221	1,2	√	SRP40				
HSPA9B	heat shock 70kDa protein 9B (mortalin-2)	73681	3069	3313			SSC1				
PARN	poly(A)-specific ribonuclease (deadenylation nuclease)	73451	43445	5073		√					
DDX3	asparagine synthetase	73243	75692	440		√	DED1				
NNP50	Novel Nucleolar Protein 50	72785	109045	55153		√	SDA1				
NNP75	Novel Nucleolar Protein 75	72485	250638				DBP6				
DDX17/P72	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 17, 72kDa	72371	349121	10521		√	DBP2				
HNRPR	heterogeneous nuclear ribonucleoprotein R	70943	15265	10236		*	NPL3				
HSPA5	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	70931	75410	3309			KAR2				
HSPA8	heat shock 70kDa protein 8	70898	180414	3312			SSA2				
NNP105	Novel Nucleolar Protein 105	70222	30670	55813		*	UTP6				
HSPA2	heat shock 70kDa protein 2	70021	75452	3306			SSA3				
HSPA1A	heat shock 70kDa protein 1B	69868	274402	3304			SSA2				
G22P1	thyroid autoantigen 70kDa (Ku antigen)	69843	197345	2547		√	YKU70				
RBM14	RNA binding motif protein 14	69492	11170	10432	2	√	HKR1				
NNP115	Novel Nucleolar Protein 115	69293	184519	65095		*	KRI1				
DDX5	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68kDa)	69148	76053	1655		√	DBP2				
NNP24/BING4	Novel Nucleolar Protein 24	68041	17930	9277	*	√	UTP7				
PES1	pescadillo homolog 1, containing BRCT domain (zebrafish)	68003	13501	23481	1*	√	NOP7				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
LMNB1	lamin B1	66408	89497	4001		√	USO1				
NNP101	Novel Nucleolar Protein 101	66302	32922	55602		*	SRP40				
NOP56	nucleolar protein 5A (56kDa with KKE/D repeat)	66236	296585	10528	1p,2	√	SIK1				
FTSJ3	FtsJ homolog 3 (E. coli)	65728	257486	117246		*	SPB1				
SENP3	sentrin/SUMO-specific protease 3	64857	118926	26168			ULP1				
HNRPL	heterogeneous nuclear ribonucleoprotein L	64087	2730	3191		*					
NNP47	Novel Nucleolar Protein 47	63569	279923	26354		√	NUG1				
DED	apoptosis antagonizing transcription factor	63133	16178	26574		√	BFR2				
COILIN	coilin	62608	966	8161			SRP40				
NOH61	putative nucleolar RNA helicase	61590	10098	54606	1*	√	DBP9				
HSPD1	heat shock 60kDa protein 1 (chaperonin)	61025	79037	3329		√	HSP60				
CCT3	chaperonin containing TCP1, subunit 3 (gamma)	60403	1708	7203			CCT3				
NOP5/NOP58	nucleolar protein NOP5/NOP58	59578	119908	51602	2	√	NOP58				
NNP114	Novel Nucleolar Protein 114	59558	151001	84916		*	UTP4				
NNP84	Novel Nucleolar Protein 84	59380	145696	9584		*	NSR1				
NNP44	Novel Nucleolar Protein 44	59103	6153	51096		√	UTP18				
DKC1	dyskeratosis congenita 1, dyskerin	58735	4747	1736	1,2*	√	CBF5				
NNP117	Novel Nucleolar Protein 117	58468	95196	79050		*	UTP19				
NNP46	Novel Nucleolar Protein 46	58132	85963	26156	+	√	YKR060W				
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	57956	25			*	ATP2				
KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	57862	159557	3838	2p	√	SRP1				
NNP45	Novel Nucleolar Protein 45	57508	2471	9933			PUF6				
nPTB	polypyrimidine tract binding protein 1	57221	172550	5725	1	√	MSL1				
P4HB	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)	57116	75655	5034			PDI1				
GRP58	glucose regulated protein, 58kDa	56782	289101	2923			PDI1				
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	56560	355912	506		√	ATP2				
NNP42	Novel Nucleolar Protein 42	55828	172589	11137	*	√	PWP1				
NNP41	Novel Nucleolar Protein 41	55213	99423			√	ROK1				
NMP200	nuclear matrix protein NMP200 related to splicing factor PRP19	55181	173980	27339	+	√	TAF5				
ASE-1	CD3-epsilon-associated protein; antisense to ERCC-1	54986	211956	10849	1	*	DEF1				
NNP39	Novel Nucleolar Protein 39	54558	322901	57050	*	√	SAS10				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NNP38	Novel Nucleolar Protein 38	54417	372421	29997	2*	√	YPL146C				
NONO	non-POU domain containing, octamer-binding	54232	355861	4841			PAB1				
NNP40	Novel Nucleolar Protein 40	54226	143187	54555		√	DBP8				
VIM	vimentin	53652	297753	7431		*	USO1				
KRT8	keratin 8	53562	242463	3856		√	USO1				
U2AF65	U2 small nuclear ribonucleoprotein auxiliary factor (65kD)	53501	7655	11338		√	RNA15				
NNP116	Novel Nucleolar Protein 116	53419	85570	54475		*	YCR072C				
NNP88	Novel Nucleolar Protein 88	53355	99969	2521		*	DDR48				
NNP123	Novel Nucleolar Protein 123	53199	50441	51067		*	MSY1				
NNP37	Novel Nucleolar Protein 37	53194	14468	56342	*	√	SSF1				
NOP52	DNA segment on chromosome 21 (unique) 2056 expressed sequence	52839	110757	8568	2p	√	RRP1				
U3-55K	U3 snoRNP-associated 55-kDa protein	51841	153768	9136	1,2p	√	RRP9				
NNP35	Novel Nucleolar Protein 35	51480	5158	23378		√	RRP8				
NNP34	Novel Nucleolar Protein 34	51388	273344	25879		√	SOF1				
TUBG1	tubulin, gamma 1	51170	21635	7283			TUB4				
EIF2S3	eukaryotic translation initiation factor 2, subunit 3 (gamma, 52kD)	51110	211539	1968		*	GCD11				
HNRPK	heterogeneous nuclear ribonucleoprotein K	50976	129548	3190		√	PBP2				
NNP33	Novel Nucleolar Protein 33	50647	284288	51202			RRP3				
TUBA6	tubulin, alpha 3	50152	272897	7846		√	TUB1				
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	50127	181165	1915		*	TEF1				
TUBB1	hypothetical protein DKFZp434N0650	49759	179661	84221		√	TUB2				
TUFM	Tu translation elongation factor, mitochondrial	49542	12084	7284		√	TUF1				
NNP36	Novel Nucleolar Protein 36	49419	218842	83743		√	RRB1				
NNP32	Novel Nucleolar Protein 32	49398	8768	55752			CDC3				
HNRPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	49229	245710	3187		√					
NNP110	Novel Nucleolar Protein 110	49086	74899	64318		*	NOC3				
DDXL	HLA-B associated transcript 1	49078	55296	7919		√	SUB2				
ETF1	eukaryotic translation termination factor 1	49031	77324	2107			SUP45				
CDC10	CDC10 cell division cycle 10 homolog (S. cerevisiae)	48787	184326	989		*	CDC3				
CSTF1	cleavage stimulation factor, 3' pre-RNA, subunit 1, 50kDa	48358	172865	1477			PWP2				
KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1	48227	119537	10657		*	MSL5				
P5	protein disulfide isomerase-related protein	48121	182429	10130			MPD1				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
KRT17	keratin 17	48106	2785	3872		*	USO1				
NNP31	Novel Nucleolar Protein 31	48063	322478	23029		√	NOP12				
KRT18	keratin 18	48032	65114	3875		*	USO1				
GRSF1	G-rich RNA sequence binding factor 1	47999	309763	2926			GBP2				
RPL4	ribosomal protein L4	47697	286	6124	r	√	RPL4B				
NNP30	Novel Nucleolar Protein 30	47694	73291	55759		√	YTM1				
EEF1A2	eukaryotic translation elongation factor 1 alpha 2	47456	2642	1917		*	TEF1				
NNP29	Novel Nucleolar Protein 29	47435	325321	57418		√	MDV1				
ACTR3	ARP3 actin-related protein 3 homolog (yeast)	47371	5321	10096			ARP3				
NNP113	Novel Nucleolar Protein 113	47260	24884	64425		*	RPA49				
PMSCL1	polymyositis/scleroderma autoantigen 1, 75kDa	46978	91728	5393	1,2p	√	RRP45				
IF4N	KIAA0111 gene product	46871	79768	9775		√	FAL1				
NNP28	Novel Nucleolar Protein 28	46672	235376	57109		√	REX4				
NNP94	Novel Nucleolar Protein 94	46486	7194	51631		*	LUC7				
SERPINH2	serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 2	46441	9930	872		*					
HNRPDL	heterogeneous nuclear ribonucleoprotein D-like	46438	170311	9987		*	HRP1				
EIF4A1	eukaryotic translation initiation factor 4A, isoform 1	46154	129673	1973		√	TIF1				
RPL3	ribosomal protein L3	46109	119598	6122	r	√	RPL3				
CSNK2A1	casein kinase 2, alpha 1 polypeptide	45144	155140	1457		*	CKA1				
ACTR2	ARP2 actin-related protein 2 homolog (yeast)	44761	42915	10097			ARP2				
TARDBP	TAR DNA binding protein	44740	193989	23435		*	HRP1				
ILF2	interleukin enhancer binding factor 2, 45kDa	44697	75117	3608		√					
CHC1	chromosome condensation 1	44057	84746	1104		*	SRM1				
METAP1	methionyl aminopeptidase 1	44046	82007	23173		*	MAP1				
NNP82	Novel Nucleolar Protein 82	43801	224137	51490			YGR283C				
PA2G4	Human erbB3 binding protein EBP1 mRNA, complete cds	43787	343258				MAP2				
NNP1	Novel Nucleolar Protein 1	43746	375025	85406							
NNP25	Novel Nucleolar Protein 25	43634	71040	55646		√	YCR087C-A				
HRB2	HIV-1 rev binding protein 2	43593	154762	11103	2		KRR1				
RBMX	RNA binding motif protein, X chromosome	42404	146381	27316		√	HRP1				
NNP23	Novel Nucleolar Protein 23	42286	3487	54663		√	NSA1				
ACTA1	actin, alpha 1, skeletal muscle	42051	1288	58		*	ACT1				
PSP1	paraspeckle protein 1	41739	16364	55269	*		PAB1				
ACTB	actin, beta	41737	288061	60		√	ACT1				
ACTG1	actin, gamma 1	41662	14376	71		*	ACT1				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NNP91	Novel Nucleolar Protein 91	41487	155595	4735		*	CDC3				
NNP81	Novel Nucleolar Protein 81	41450	274430	6838	2		YKL082C				
NNP3	Novel Nucleolar Protein 3	41401	38114	55299		√	BRX1				
NNP26	Novel Nucleolar Protein 26	41193	71827	23212	2	√	RRS1				
RBM4	RNA binding motif protein 4	40940	42826	5936		√	PAB1				
RNAC	RNA cyclase homolog	40871	113052	10171	2	√	RCL1				
RFC3	replication factor C (activator 1) 3 (38kD)	40556	115474	5983		*	RFC5				
NNP22	Novel Nucleolar Protein 22	40081	287863	80135	2p	√	RPF1				
HNRPA3	heterogeneous nuclear ribonucleoprotein A3	39686	249247			*	HRP1				
H2AFY	H2A histone family, member Y	39601	75258	9555		√	HTA2				
RPA40	RNA polymerase I subunit	39250	5409	9533	1*	√	RPC40				
RFC2	replication factor C (activator 1) 2, 40kDa	39157	139226	5982			RFC4				
HNRPA1	heterogeneous nuclear ribonucleoprotein A1	38846	249495	3178		√	HRP1				
KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	38800	13565	10656	1	*	MSL5				
ANXA2	annexin A2	38604	217493	302		√					
NNP20	Novel Nucleolar Protein 20	37884	59425	79707							
PCBP1	poly(rC) binding protein 1	37526	2853	5093			PBP2				
PPP1CA	protein phosphatase 1, catalytic subunit, gamma isoform	37512	79081	5501		√	GLC7				
NNP126	Novel Nucleolar Protein 126	37432	55896			*	RPA43				
ASNA1	arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)	37204	165439	439	1	*	ARR4				
PPP1CB	protein phosphatase 1, catalytic subunit, beta isoform	37187	21537	5500		√	GLC7				
NNP19	Novel Nucleolar Protein 19	37035	99829	54984			PXR1				
RP1	microtubule-associated protein, RP/EB family, member 2	37031	78335	10982			BIM1				
HNRPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	36926	279681	3189			NPL3				
GAPD	glyceraldehyde-3-phosphate dehydrogenase	36054	356771	2597		*	TDH3				
HNRPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	36006	232400	3181		√	HRP1				
HNRPAB	heterogeneous nuclear ribonucleoprotein A/B	35968	81361	3182		√	HRP1				
NNP96	Novel Nucleolar Protein 96	35711	9043	25983		*	LCP5				
MDH2	malate dehydrogenase 2, NAD (mitochondrial)	35503	343521	4191		*	MDH1				
APEX	APEX nuclease (multifunctional DNA repair enzyme)	35432	73722	328		*					
NNP78	Novel Nucleolar Protein 78	35397	77135	84549		√	MAK16				
NNP16	Novel Nucleolar Protein 16	35369	301947	94009							
BYSL	bystin-like	34864	106880	705			ENP1				
NNP17	Novel Nucleolar Protein 17	34820	346868	10969	2	√	EBP2				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
RPP40	ribonuclease P, 40kD subunit	34563	115823	10799		*					
RPL5	ribosomal protein L5	34448	372205		r	√	RPL5				
RPLP0	ribosomal protein, large, P0	34274	350108	6175	r		RPP0				
NNP18	Novel Nucleolar Protein 18	34253	142838	84365			NOP15				
FBL	fibrillarin	33784	99853	2091	1,2*	√	NOP1				
NNP14	Novel Nucleolar Protein 14	33757	91579	92856			IMP4				
HNRPC	heterogeneous nuclear ribonucleoprotein C (C1/C2)	33688	182447	3183		√	NAB3				
SFRS10	splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)	33666	30035	6434		*	NSR1				
LAMR1	laminin receptor 1 (ribosomal protein SA, 67kDa)	32854	356264	3921	r	√	RPS0A				
SLC25A5	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5	32852	79172	292		*	AAC3				
NNP10	Novel Nucleolar Protein 10	32789	211973	23404		√	RRP4				
RPL6	ribosomal protein L6	32728	349961	6128	r	√	RPL6B				
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	32603	9614	4869		√	SIS2				
NNP90	Novel Nucleolar Protein 90	32334	239934	27341		*	RRP7				
NNP120	Novel Nucleolar Protein 120	32321	100134	84135		*	UTP15				
NNP15	Novel Nucleolar Protein 15	31835	182877	23016		√	RRP45				
NNP121	Novel Nucleolar Protein 121	31812	39504	84319		*					
SFRS5	splicing factor, arginine/serine-rich 5	31264	166975	6430		√	HRB1				
ABT1	activator of basal transcription 1	31079	109428	29777	2p	√	YNR054C				
DNAJC8	DnaJ (Hsp40) homolog, subfamily C, member 8	30987	74711	22826			HLJ1				
HNRPA0	heterogeneous nuclear ribonucleoprotein A0	30901	77492	10949		*	HRP1				
NNP13	Novel Nucleolar Protein 13	30448	111449	51118		√	UTP11				
NNP12	Novel Nucleolar Protein 12	30065	8170	10412		√	NSA2				
NNP11	Novel Nucleolar Protein 11	30040	375585	11340	2p	√	RRP45				
RPS3A	ribosomal protein S3A	30007	77039	6189	r	√	RPS1B				
RPL7A	ribosomal protein L7a	29996	99858	6130	r	√	RPL8B				
PHB	prohibitin	29804	75323	5245			PHB1				
RPS4X	ribosomal protein S4, X-linked	29598	108124	6191	r*	√	RPS4A				
RRP40	exosome component Rrp40	29572	177677	51010	2p	√	RRP40				
RARG-1	retinoic acid repressible protein	29426	106346	51406		√	BUD21				
RPP30	ribonuclease P (30kD)	29321	139120	10556		*	RPP1				
NNP125	Novel Nucleolar Protein 125	29293	375009			*	RPF2				
RPL7	ribosomal protein L7	29226	356593	6129	r	√	RPL7B				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NNP80	Novel Nucleolar Protein 80	28973	56043	51018		√	YPR143W				
RPS6	ribosomal protein S6	28639	350166	6194	r*		RPS6A				
NNP9	Novel Nucleolar Protein 9	28533	374582			√	RPL7A				
SNRPA1	small nuclear ribonucleoprotein polypeptide A'	28444	80506	6627		√	LEA1				
NNP74	Novel Nucleolar Protein 74	28235	351784	118460			SKI6				
RPL8	ribosomal protein L8	28025	178551	6132	r	√	RPL2B				
NNP108	Novel Nucleolar Protein 108	27924	193384	56902		*	YOR145C				
U2AF1	U2(RNU2) small nuclear RNA auxillary factor 1	27872	59271	7307		*					
SFRS1	splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)	27745	73737	6426	2	√	NPL3				
NNP129/RPS0	Novel Nucleolar Protein 129	27505	274201	51154	r		MRT4				
SFRS7	splicing factor, arginine/serine-rich 7, 35kDa	27367	184167	6432		√	IST3				
ITGB4BP	integrin beta 4 binding protein	26599	5215	3692		√	TIF6				
NNP8	Novel Nucleolar Protein 8	26383	343589	54512	2p	√	SKI6				
NNP7	Novel Nucleolar Protein 7	26263	12045	10436		√	EMG1				
CPSF5	cleavage and polyadenylation specific factor 5, 25 kD subunit	26227	9605	11051		*					
NNP6	Novel Nucleolar Protein 6	25981	241520	10189			NSR1				
UBB	ubiquitin B	25762	183842	7314		√	UBI4				
SFRS2	splicing factor, arginine/serine-rich 2	25575	73965	6427		√	NSR1				
SFRS9	splicing factor, arginine/serine-rich 9	25542	77608	8683	2		NPL3				
NNP21-new	Novel Nucleolar Protein 21	25456	126522	85395							
POP4	POP4 (processing of precursor , S. cerevisiae) homolog	25425	82238	10775	1	*	POP4				
RRP46	exosome component Rrp46	25249	283741	56915		*	RRP46				
HMG1	high-mobility group (nonhistone chromosomal) protein 1	24993	337757	3146		*	NHP6A				
CSNK2B	casein kinase 2, beta polypeptide	24916	165843	1460		*	CKB2				
RPL10A	ribosomal protein L10a	24831	334895	4736	r		RPL1A				
NNP5	Novel Nucleolar Protein 5	24663	326372	79159			YDR412W				
POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	24612	24301	5434		√	RPB5				
RPL13	ribosomal protein L13	24292	356148		r	√	RPL13B				
RPS8	ribosomal protein S8	24205	151604	6202	r	√	RPS8B				
RPL15	calcium binding protein P22	24146	85301	11261	r	√	RPL15B				
RPL14	ribosomal protein L14	23803	738	9045	r	√	RPL14B				
RPL13A	ribosomal protein L13A	23577	356678		r	√	RPL16A				
RPL19	ribosomal protein L19	23466	252723	6143	r	*	RPL19A				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NNP100	Novel Nucleolar Protein 100	23370	343173	51077		*	YDR339C				
RPS5	ribosomal protein S5	22876	76194	6193	r	√	RPS5				
PPIB	peptidylprolyl isomerase B (cyclophilin B)	22742	699	5479		√	CPR5				
RPS9	ribosomal protein S9	22591	180920	6203	r	√	RPS9B				
H1F5	H1 histone family, member 5	22580	131956	3009		*	HHO1				
NOLA1	nucleolar protein family A, member 1 (H/ACA small nucleolar RNPs)	22348	69851	54433	2	√	GAR1				
CBX5	chromobox homolog 5 (HP1 alpha homolog, Drosophila)	22225	89232	23468		*					
TASR1	TLS-associated serine-arginine protein 1	22222	355805				NSR1				
RPS7	ribosomal protein S7	22127	301547	6201	1r	*	RPS7A				
PPIF	peptidylprolyl isomerase F (cyclophilin F)	22040	173125	10105		√	CPR1				
RPL9	ribosomal protein L9	21863	157850	6133	r	√	RPL9B				
MRPS4	chromosome 15 open reading frame 12	21850	6118	55272	2p	√	IMP3				
RPL18	ribosomal protein L18	21634	343354	6141	r	√	RPL18A				
CSL4	exosomal core protein CSL4	21452	14415	51013		*	CSL4				
CBX1	chromobox homolog 1 (HP1 beta homolog Drosophila)	21418	77254	10951			MNN4				
RPL17	ribosomal protein L17	21397	82202	6139	r	√	RPL17A				
H1F2	H1 histone family, member 2	21365	7644	3006		*	HHO1				
NNP4	Novel Nucleolar Protein 4	21148	279918	51491		√					
HP1g	chromobox homolog 3 (HP1 gamma homolog, Drosophila)	20823	278554	11335	*	√					
RPL18A	ribosomal protein L18a	20762	163593	6142	r	*	RPL20B				
CSRP1	cysteine and glycine-rich protein 1	20567	108080	1465							
ARPC3	actin related protein 2/3 complex, subunit 3, 21kDa	20547	6895	10094		√	ARC18				
NNP98	Novel Nucleolar Protein 98	20463	268049	51388		*	NIP7				
RPL11	ribosomal protein L11	20252	179943	6135	r	√	RPL11A				
NNP107	Novel Nucleolar Protein 107	19621	284162	51187		*	RLP24				
TPT1	tumor protein, translationally-controlled 1	19595	279860	7178	+	√	YKL056C				
NPM3	nucleophosmin/nucleoplasmin, 3	19344	90691	10360		√					
SFRS3	splicing factor, arginine/serine-rich 3	19330	167460	6428	2	√					
RPS11	ribosomal protein S11	18431	182740	6205	r	√	RPS11A				
PPIA	peptidylprolyl isomerase A (cyclophilin A)	18012	342389	5478		√	CPR1				
RPL12	ribosomal protein L12	17819	356318		r	√	RPL12A				
RPS18	ribosomal protein S18	17719	275865	6222	r	√	RPS18B				
RPL23A	ribosomal protein L23a	17695	350046	6147	r	√	RPL25				
NNP85	Novel Nucleolar Protein 85	17657	356485			*	RPL21A				
NNP83	Novel Nucleolar Protein 83	17445	356355			*	RPS31				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
RPL26	ribosomal protein L26	17258	91379	6154	r	√	RPL26B				
RPS13	ribosomal protein S13	17222	165590	6207	r	√	RPS13				
NOLA2	nucleolar protein family A, member 2 (H/ACA small nucleolar RNPs)	17201	23990	55651	2	√	NHP2				
RBM3	RNA binding motif protein 3	17170	301404	5935		*	NSR1				
POLR2H	polymerase (RNA) II (DNA directed) polypeptide H	17143	3128	5437			RPB8				
NNP93	Novel Nucleolar Protein 93	16955	22393	8562		*	YJR014W				
EIF5A	eukaryotic translation initiation factor 5A	16832	119140	1984		√	HYP2				
RPL27A	ribosomal protein L27A	16561	356342		r	√	RPL28				
RPS16	ribosomal protein S16	16445	80617	6217	r	√	RPS16A				
RPS14	ribosomal protein S14	16273	244621	6208	r	*	RPS14B				
RPL32	ribosomal protein L32	15860	169793	6161	r	*	RPL32				
RPL27	ribosomal protein L27	15798	111611	6155	r*	√	RPL27B				
RPL28	ribosomal protein L28	15748	4437	6158	r	*					
NNP2	Novel Nucleolar Protein 2	15434	NA								
NNP127	Novel Nucleolar Protein 127	15257	348668	126961		*	HHT2				
NNP128	Novel Nucleolar Protein 128	15036	322802			*	RPL23B				
NNP106	Novel Nucleolar Protein 106	14960	9676	55856		*					
RPL22	ribosomal protein L22	14787	326249	6146	r	√	RPL22A				
RPL35	ribosomal protein L35	14551	182825	11224	r	*	RPL35A				
SRP14	signal recognition particle 14kD	14544	355573								
RPS12	ribosomal protein S12	14526	356711		r	√	RPS12				
RPL31	ribosomal protein L31	14463	184014	6160	r	√	RPL31B				
PC4	activated RNA polymerase II transcription cofactor 4	14365	356473			*	SUB1				
NNP99	Novel Nucleolar Protein 99	14199	79259	51504		*	YNR046W				
NHP2L1	NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)	14174	182255	4809	1,2*	√	SNU13				
H2AFO	H2A histone family, member O	14095	795	8337		*	HTA1				
H2AFE	H2A histone family, member E	13936	274590			√	HTA1				
H2BFQ	H2B histone family, member Q	13920	2178	8349		*	HTB1				
SNRPD3	small nuclear ribonucleoprotein D3 polypeptide 18kDa	13916	1575	6634	*	√	SMD3				
H2BFA	H2B histone family, member L	13906	356901	8347	*	√	HTB1				
H2BFG	H2B histone family, member G	13775	182137	8343		*	HTB1				
H2AFZ	H2A histone family, member Z	13553	119192	3015		*	HTZ1				
SNRPD2	small nuclear ribonucleoprotein D2 polypeptide (16.5kD)	13527	53125	6633	1*	*	SMD2				
RPL34	ribosomal protein L34	13293	250895	6164	r	*	RPL34B				

Table III-3 Nucleolar Proteome from current studies

Gene Name	Description	calc. Mr	UniGene	LocusID	Loc	LCMS	Yeast GN	Sc	Sp	Ce	Dm
NNP89	Novel Nucleolar Protein 89	12784	356255	23299		*	RPL30				
RPL35A	ribosomal protein L35a	12538	287361	6165	r	*	RPL33B				
NNP92	Novel Nucleolar Protein 92	12259	118757	2079		*					
RPL36	ribosomal protein L36	12235	343443	25873	r	*	RPL36B				
HCS	cytochrome c	11618	169248	54205		*	CYC1				
H4F2	H4 histone family, member H	11367	93758	8365		√	HHF2				
SRP9	signal recognition particle 9kD	9981	75975	6726		*					
RPS27L	ribosomal protein S27-like	9478	108957	51065	r	*	RPS27B				
NOLA3	nucleolar protein family A, member 3 (H/ACA small nucleolar RNPs)	7706	14317	55505		*	NOP10				

Table III-3 **Nucleolar Proteome resulting from current studies**. Proteins are listed in descending order according to their calculated molecular weight. "Unigene" and "LocusID" indicate the cDNA cluster and genomic loci respectively assigned by NCBI respectively. "Code" indicates the cellular localization (if known), where r = ribosomal, 1 = published nucleolar localization for the human protein, 2 = published nucleolar localization for protein homologues from other species, * = tagged in this study and shown to be nucleolar, and + = tagged in this study and not observed to be nucleolar. "LCMS" indicates whether the proteins were identified from LCMS, where √ = LCMS only, * = LCMS and 1D/2D study and "empty" = 1D/2D only. The colours indicated at the left of each entry relate to the classification of each protein as assigned in Table III-2. 'Yeast GN' shows the closest match of a yeast gene product with the corresponding human protein. The shaded panels represent the results of BLASTp searches for each protein against species-specific proteome databases from EBI/SGD databases: *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Drosophila melanogaster* (Dm) and *Caenorhabditis elegans* (Ce). The results are shaded such that a black panel represents the expectancy value (e-value) of 0.0 and thus indicates a nearly perfect match; 80% grey for e-value < $1e^{-100}$; 50% for $1e^{-100}$ < e-value < $1e^{-75}$; 40% for $1e^{-75}$ < e-value < $1e^{-50}$; 25% for $1e^{-50}$ < e-value < $1e^{-25}$ and a white panel for e-value > $1e^{-25}$.

Surprisingly, the subsequent addition of 136 newly identified nucleolar proteins from the LC-MS/MS data resulted in minimal change to the distribution of proteins between categories as previously assigned (Figure III-3). From the knowledge of the functions of orthologues in other organisms, particularly in budding yeast, some of the proteins were then assigned to be related to ribosomal biogenesis and denoted in green in Figure III-3.

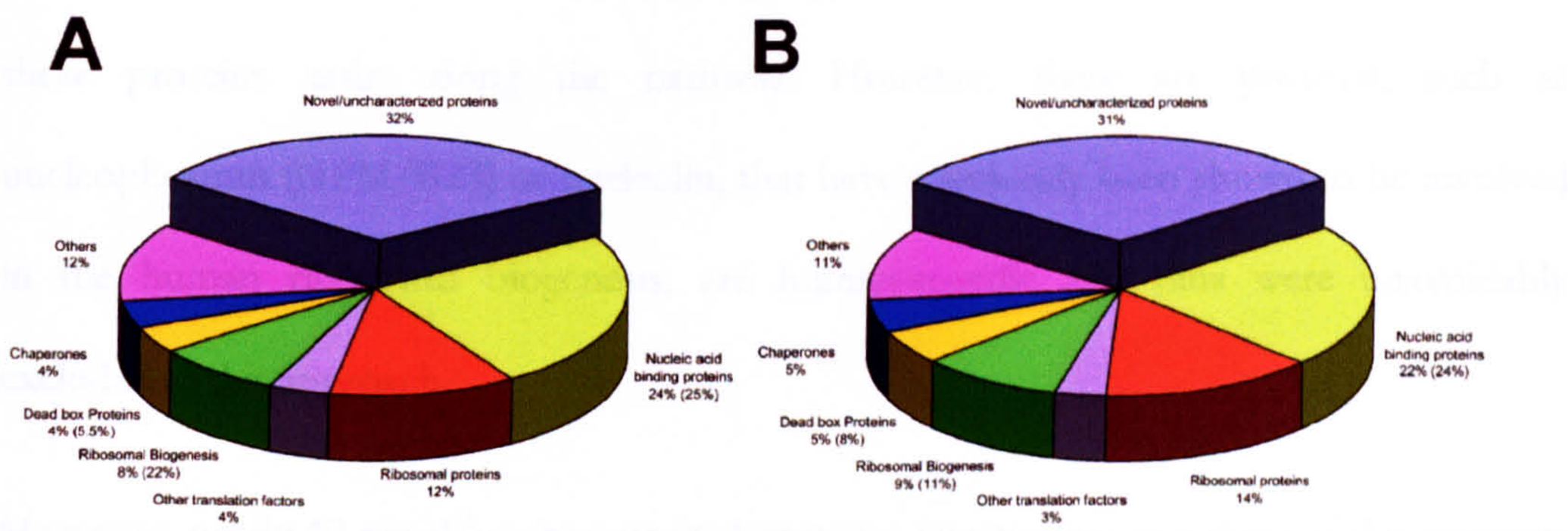


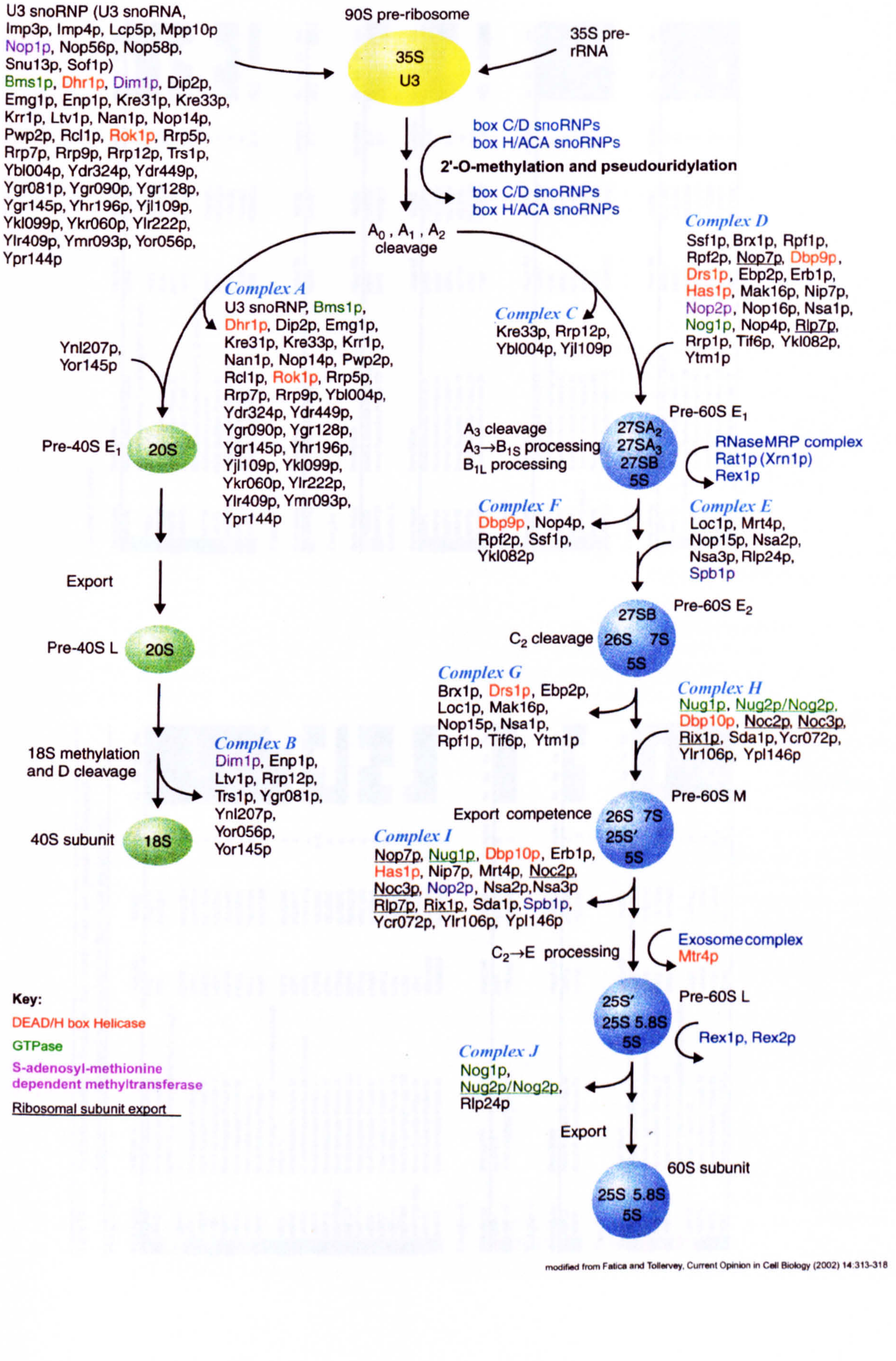
Figure III-3 **Distribution of conserved motifs and putative functional categories of the identified proteins in the nucleolar proteome.** (A) Current analysis and (B) 1D/2D gel analysis data (Andersen et al, 2002). The percentages in brackets indicate the maximum number of proteins that could possibly be classified in the category in question. For example, 22% of the proteins may be related to ribosomal biogenesis from the current analysis (A), due to their yeast orthologues being present in the ribosomal synthesis pathway.

The ribosomal biogenesis pathway was so far studied mainly in yeast, while its characterisation in other eukaryotes, especially humans, lag far behind (Fatica and Tollervey, 2002). The accumulated knowledge of genetic defects in the yeast ribosome synthesis pathway during the last decade as well as the recent success in the identification of pre-ribosomes using mass spectrometry has expanded our understanding of the ribosome synthesis pathway dramatically (Figure III-4 and http://www.proteome.com). Clear human homologues identified from our analysis could be found in the pathway as illustrated in Table III-4. In this way, our nucleolar proteome can annotate part of the human genome to be functionally related to ribosomal biogenesis. Future analysis of these human nucleolar proteins is likely to improve our understanding of human ribosomal synthesis pathway.

Drawing on published data, as well as identifying homologues from the yeast ribosome biogenesis pathway as shown in Figure III-4, we have found 92 homologues (Table III-4). Table III-4 shows all the possible 92 human homologues in the nucleolar proteome that could potentially be involved in ribosomal biogenesis. Proteins previously proven to be involved in the human ribosomal biogenesis such as POP4, FIB and NOP52 were shown to have similar functions as in yeast homologues and were organised according to how these proteins assist along the pathway. However, there are proteins, such as nucleophosmin (NPM/B23) or nucleolin, that have previously been shown to be involved in the human ribosomal biogenesis, are human-specific and thus were unavoidably excluded in this approach.

Moreover, nearly 40 novel/uncharacterised proteins, constituting one-third of this category in the nucleolar proteome, could potentially have functions related to ribosomal biogenesis based on the function of their yeast homologues (Figure III-4). For other human proteins with limited characterisation, such as Bystin-like protein BYSL and many DEAD/H box helicases, we can now more precisely pinpoint their possible functions. This illustrates the potential for using the information from this study to further elucidate the human ribosomal biogenesis in the near future.

Figure III-4 **Ribosomal biogenesis pathway in yeast.** The 90S pre-ribosomal complex is proposed to contain the 35S rRNA and the U3 snoRNA. The early pre-rRNA cleavages at sites A_0 to A_2 lead to the separation of the pre-40S and pre-60S particles. In both pathways a series of predicted intermediates are drawn, which are designated early (E), middle (M) and late (L) according to their positions on the proposed pathway. The processing steps envisaged to be associated with each of these complexes are indicated, as is the likely time of export to the cytoplasm. Note that it is very probable that other pre-ribosomal complexes exist in addition to those shown, and it is not clear in what order the components are gained and lost between the complexes. The pre-60S pathway includes only proteins for which corroborating data exist supporting a direct role in ribosome synthesis. This figure is modified from Fatica and Tollervey, 2002.



modified from Fatica and Tollervey, Current Opinion in Cell Biology (2002) 14:313-318

Table III-4 **Nucleolar proteins found in this study that are putative ribosomal biogenesis factors based on homology to their yeast counterparts**

Complex A

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP69	Novel Nucleolar Protein 69	BMS1	YPL217C	A				
NNP63	Novel Nucleolar Protein 63	DIP2	YLR129W	A				
DDX37	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 37	ECM16	YMR128W	A				
NNP7	Novel Nucleolar Protein 7	EMG1	YLR186W	A				
NNP66	Novel Nucleolar Protein 66	KRE33	YNL132W	A				
HRB2	HIV-1 rev binding protein 2	KRR1	YCL059C	A				
NNP87	Novel Nucleolar Protein 87	NOP14	YDL148C	A				
NNP62	Novel Nucleolar Protein 62	PWP2	YCR057C	A				
RNAC	RNA cyclase homolog	RCL1	YOL010W	A				
NNP41	Novel Nucleolar Protein 41	ROK1	YGL171W	A				
RRP5	likely ortholog of mouse programmed cell death protein 11	RRP5	YMR229C	A				
NNP90	Novel Nucleolar Protein 90	RRP7	YCL031C	A				
U3-55K	U3 snoRNP-associated 55-kDa protein	RRP9	YPR137W	A				
NNP71	Novel Nucleolar Protein 71	UTP10	YJL109C	A, C				
NNP13	Novel Nucleolar Protein 13	UTP11	YKL099C	A				
NNP43/SAZD	Novel Nucleolar Protein 43	UTP13	YLR222C	A				
NNP120	Novel Nucleolar Protein 120	UTP15	YMR093W	A				
NNP117	Novel Nucleolar Protein 117	UTP19	YPR144C	A				
NNP49	Novel Nucleolar Protein 49	UTP21	YLR409C	A				
NNP27	Novel Nucleolar Protein 27	UTP22	YGR090W	A				
NNP114	Novel Nucleolar Protein 114	UTP4	YDR324C	A				
NNP105	Novel Nucleolar Protein 105	UTP6	YDR449C	A				
NNP24/BING4	Novel Nucleolar Protein 24	UTP7	YER082C	A				
NNP52	Novel Nucleolar Protein 52	YGR145W	YGR145W	A				
NNP46	Novel Nucleolar Protein 46	YKR060W	YKR060W	A				

Complex B

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
BYSL	bystin-like	ENP1	YBR247C	B				
NNP108	Novel Nucleolar Protein 108	YOR145C	YOR145C	B				

Complex C

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP68	Novel Nucleolar Protein 68	RRP12	YPL012W	C				
NNP71	Novel Nucleolar Protein 71	UTP10	YJL109C	A, C				

Complex D

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP3	Novel Nucleolar Protein 3	BRX1	YOL077C	D, G				
NOH61	putative nucleolar RNA helicase	DBP9	YLR276C	D, F				
NNP17	Novel Nucleolar Protein 17	EBP2	YKL172W	D, G				
NNP54/BOP1	Novel Nucleolar Protein 54	ERB1	YMR049C	D				
DDX18	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated)	HAS1	YMR290C	D				
NNP78	Novel Nucleolar Protein 78	MAK16	YAL025C	D				
NNP51	Novel Nucleolar Protein 51	NOG1	YPL093W	D, J				
NOL1	nucleolar protein 1, 120kDa	NOP2	YNL061W	D				

Complex D (*cont'd*)

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP59	Novel Nucleolar Protein 59	NOP4	YPL043W	D				
PES1	pescadillo homolog 1, containing BRCT domain (zebrafish)	NOP7	YGR103W	D				
NNP23	Novel Nucleolar Protein 23	NSA1	YGL111W	D, G				
NNP22	Novel Nucleolar Protein 22	RPF1	YHR088W	D				
NNP125	Novel Nucleolar Protein 125	RPF2	YKR081C	D, F				
NOP52	DNA segment on chromosome 21 (unique) 2056 expressed sequence	RRP1	YDR087C	D				
NNP37	Novel Nucleolar Protein 37	SSF1	YHR066W	D				
ITGB4BP	integrin beta 4 binding protein	TIF6	YPR016C	D				
NNP81	Novel Nucleolar Protein 81	YKL082C	YKL082C	D				
NNP30	Novel Nucleolar Protein 30	YTM1	YOR272W	D, G				

Complex E

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP107	Novel Nucleolar Protein 107	RLP24	YLR009W	E, J				

Complex F

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NOH61	putative nucleolar RNA helicase	DBP9	YLR276C	D, F				
NNP125	Novel Nucleolar Protein 125	RPF2	YKR081C	D, F				

Complex G

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP3	Novel Nucleolar Protein 3	BRX1	YOL077C	D, G				
NNP48	Novel Nucleolar Protein 48	DRS1	YLL008W	G				
NNP17	Novel Nucleolar Protein 17	EBP2	YKL172W	D, G				
NNP18	Novel Nucleolar Protein 18	NOP15	YNL110C	G				
NNP23	Novel Nucleolar Protein 23	NSA1	YGL111W	D, G				
NNP30	Novel Nucleolar Protein 30	YTM1	YOR272W	D, G				

Complex H

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP76	Novel Nucleolar Protein 76	DBP10	YDL031W	H				
MDN1	MDN1, midasin homolog (yeast)	MDN1	YLR106C	H, I				
NNP97	Novel Nucleolar Protein 97	MDN1	YLR106C	H, I				
NNP58	Novel Nucleolar Protein 58	NOC2	YOR206W	H, I				
NNP57/NGB	Novel Nucleolar Protein 57	NOG2	YNR053C	H, J				
NNP50	Novel Nucleolar Protein 50	SDA1	YGR245C	H, I				
NNP116	Novel Nucleolar Protein 116	YCR072C	YCR072C	H, I				

Complex I

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
MDN1	MDN1, midasin homolog (yeast)	MDN1	YLR106C	H, I				
NNP97	Novel Nucleolar Protein 97	MDN1	YLR106C	H, I				
NNP129/RPS0	Novel Nucleolar Protein 129	MRT4	YKL009W	I				
NNP98	Novel Nucleolar Protein 98	NIP7	YPL211W	I				
NNP58	Novel Nucleolar Protein 58	NOC2	YOR206W	H, I				
NNP110	Novel Nucleolar Protein 110	NOC3	YLR002C	I				
NNP12	Novel Nucleolar Protein 12	NSA2	YER126C	I				

Complex I

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP47	Novel Nucleolar Protein 47	NUG1	YER006W	I				
NNP50	Novel Nucleolar Protein 50	SDA1	YGR245C	H, I				
FTSJ3	FtsJ homolog 3 (E. coli)	SPB1	YCL054W	I				
NNP116	Novel Nucleolar Protein 116	YCR072C	YCR072C	H, I				
NNP86	Novel Nucleolar Protein 86	YLL034C	YLL034C	I				
NNP38	Novel Nucleolar Protein 38	YPL146C	YPL146C	I				

Complex J

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP51	Novel Nucleolar Protein 51	NOG1	YPL093W	D, J				
NNP57/NGB	Novel Nucleolar Protein 57	NOG2	YNR053C	H, J				
NNP107	Novel Nucleolar Protein 107	RLP24	YLR009W	E, J				

Exosome

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP95	Novel Nucleolar Protein 95	DIS3	YOL021C	Exosome				
NNP10	Novel Nucleolar Protein 10	RRP4	YHR069C	Exosome				
RRP40	exosome component Rrp40	RRP40	YOL142W	Exosome				
PMSCL1	polymyositis/scleroderma autoantigen 1, 75kDa	RRP45	YDR280W	Exosome				
RRP46	exosome component Rrp46	RRP46	YGR095C	Exosome				
PMSCL2	polymyositis/scleroderma autoantigen 2, 100kDa	RRP6	YOR001W	Exosome				
NNP8	Novel Nucleolar Protein 8	SKI6	YGR195W	Exosome				

H/ACA snoRNP

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
DKC1	dyskeratosis congenita 1, dyskerin	CBF5	YLR175W	HACAsnoRNP				
NOLA1	nucleolar protein family A, member 1 (H/ACA small nucleolar RNPs)	GAR1	YHR089C	HACAsnoRNP				
NOLA2	nucleolar protein family A, member 2 (H/ACA small nucleolar RNPs)	NHP2	YDL208W	HACAsnoRNP				
NOLA3	nucleolar protein family A, member 3 (H/ACA small nucleolar RNPs)	NOP10	YHR072W-A	HACAsnoRNP				

RNase MRP complex

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
NNP61	Novel Nucleolar Protein 61	POP1	YNL221C	MRP				
POP4	POP4 (processing of precursor , S. cerevisiae) homolog	POP4	YBR257W	MRP				

U3 snoRNP

Gene Name	Description	Yeast GN	SGDID	Complex	Sc	Sp	Ce	Dm
MRPS4	chromosome 15 open reading frame 12	IMP3	YHR148W	U3snoRNP				
NNP14	Novel Nucleolar Protein 14	IMP4	YNL075W	U3snoRNP				
NNP96	Novel Nucleolar Protein 96	LCP5	YER127W	U3snoRNP				
MPHOPSH10	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein)	MPP10	YJR002W	U3snoRNP				
FBL	fibrillarin	NOP1	YDL014W	U3snoRNP				
NOP5/NOP58	nucleolar protein NOP5/NOP58	NOP58	YOR310C	U3snoRNP				
NOP56	nucleolar protein 5A (56kDa with KKE/D repeat)	SIK1	YLR197W	U3snoRNP				
NHP2L1	NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)	SNU13	YEL026W	U3snoRNP				
NNP34	Novel Nucleolar Protein 34	SOF1	YLL011W	U3snoRNP				

Table III-4 **Nucleolar proteins found in this study that are putative ribosomal biogenesis factors based on homology to their yeast counterparts**. 'Yeast GN' shows the closest matching yeast gene to the corresponding protein. The shaded panels represent the results of BLASTp searches for each protein against species-specific proteome database from EBI/SGD databases: *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Drosophila melanogaster* (Dm) and *Caenorhabditis elegans* (Ce). The results are shaded such that a black panel represents the expectancy value (e-value) of 0.0 and thus indicates a nearly perfect match; 80% grey for e-value < $1e^{-100}$; 50% for $1e^{-100}$ < e-value < $1e^{-75}$; 40% for $1e^{-75}$ < e-value < $1e^{-50}$; 25% for $1e^{-50}$ < e-value < $1e^{-25}$ and a white panel for e-value > $1e^{-25}$. It should be noted that the e-value between known yeast homologues of human proteins can be as low as $1e^{-10}$.

III.4.1.3 Amino acid composition

Taking advantage of the large nucleolar proteome described in this work, I analysed whether certain amino acids are enriched, as comparing with randomly generated sequences (Figure III-5). For comparisons, I generated sets of protein sequences extracted from either the human-specific proteomes or the human, nuclear-specific proteomes and each set contains approximately the same number of sequences and amino acids as the identified nucleolar proteome. Apparently, charged amino acids such as glutamate, aspartate, lysine and arginine are more favoured, when compared with the randomly generated nuclear protein sequences, whilst neutral amino acids, such as proline and cysteine, are highly disfavoured. Surprisingly, the polar amino acids serine, threonine and tyrosine, which are important targets for phosphorylation were not common within the nucleolar protein sequences. This apparent amino acid bias may reflect nucleolar targeting

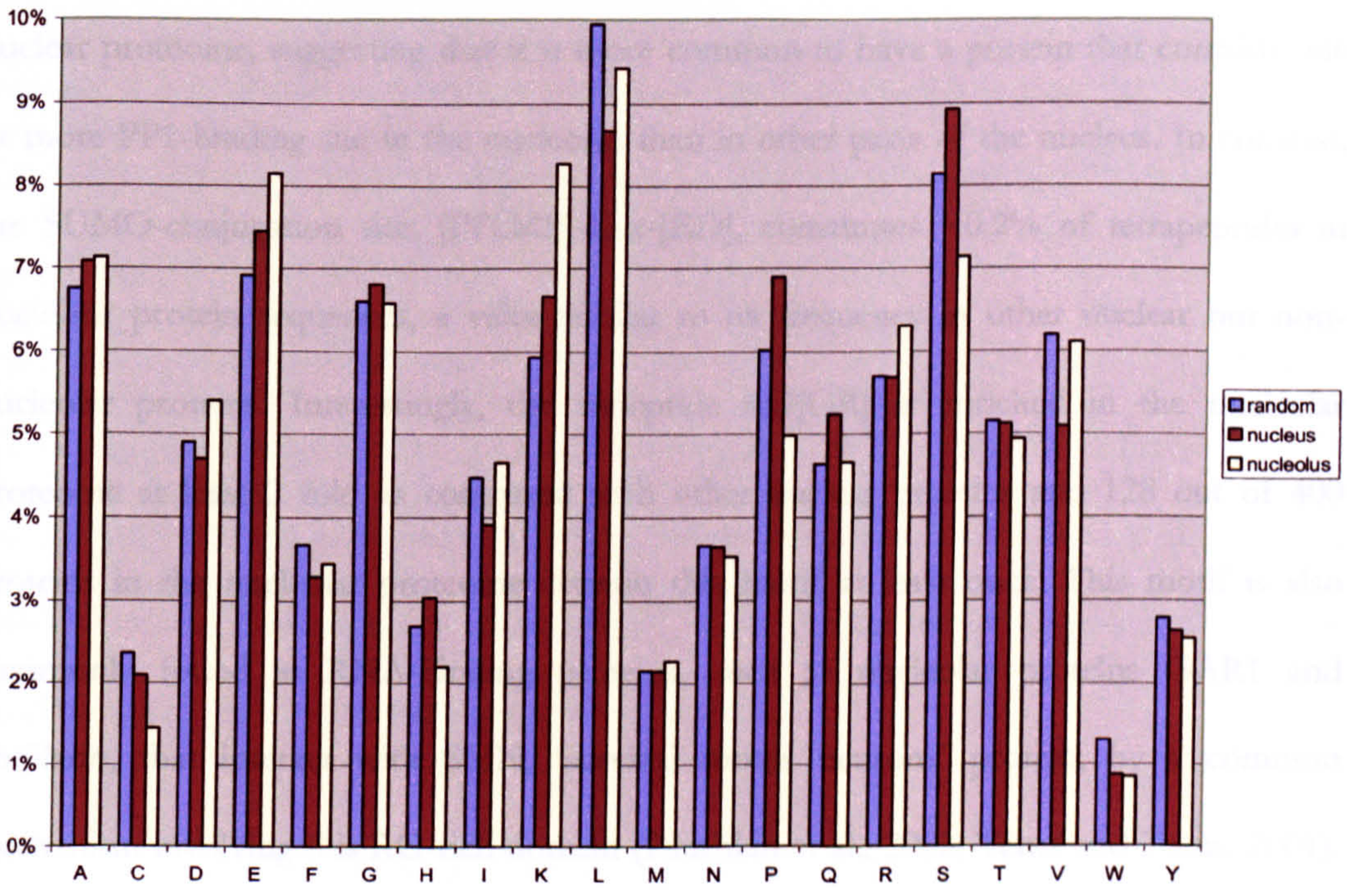


Figure III-5 Abundance of individual amino acids in the nucleolar protein sequence

motifs, although these remain poorly characterised. The initial search for such motifs using software such as MEME (Bailey and Elkan, 1994), was only sensitive enough to retrieve the motifs mentioned in Section III.4.1.2.

Whilst it was not possible to isolate new motifs, it was possible to search for other known short peptide motifs. The database was searched for all the possible tripeptides, tetrapeptides and pentapeptides within the nucleolar proteome using customised written scripts. For a protein sequence of *n* amino acids, it consists of *n-2* tripeptides, *n-3* tetrapeptides and *n-4* pentapeptides. For example, a pentapeptide sequence of MALAV has 2 tetrapeptides (MALA and ALAV) and 3 tripeptides (MAL, ALA and LAV).

Certain motifs show a specific enrichment in the human nucleolar proteome. The Protein Phosphatase 1 (PP1)-binding motif, [KR]-V-x-F, constitutes 0.032% of all tetrapeptide sequences present in the nucleolar proteome. This is ~1.4 fold more common than in the nuclear proteome, suggesting that it is more common to have a protein that contains one or more PP1-binding site in the nucleolus than in other parts of the nucleus. In contrast, the SUMO-conjugation site, [IVLMF]-K-x-[ED], constitutes ~0.2% of tetrapeptides in nucleolar protein sequences, a value similar to its frequency in other nuclear but non-nucleolar proteins. Interestingly, the tripeptide RG[GR] is enriched in the nucleolar proteome at least 2 fold as compared with other nuclear proteins and 128 out of 400 proteins in the nucleolar proteome contain this motif at least once. This motif is also commonly found in RNA-binding proteins, such as nucleolar proteins GAR1 and fibrillarin, that interact with SMN, 'survival motor neurons' protein, by a common mechanism involving this RG-rich domain (Paushkin et al., 2002; Terns and Terns, 2001). The enrichment of such motifs possibly reflects the unknown, functional properties of the human nucleolus that are distinct from other subnuclear structures.

III.4.2 Profiling

The bioinformatics analyses performed so far look at each protein or gene individually and determine whether it may contain certain motifs, or whether its orthologues in other model organisms perform certain functions. However, proteins can also be analysed either pairwise, or in clusters. For example, proteins that function in a common complex may be expressed co-ordinately in multiple organisms. This is a reasonable assumption based upon the fact that proteins rarely work alone and many pathways or complexes are crippled by the loss of individual components. Such phenomena have been observed in yeast, where co-inherited proteins are usually functionally related (Pellegrini et al., 1999).

To analyse and present the homologues from 80 completed genome sequences (the complete list of organisms is available in Table III-S2), we can view them as a profile, where each profile is a single row consisting of 80 elements that individually encode either the presence (100% black), or absence (0% black), of sequence homologues in a particular genome. Each profile is a representation of the homologues across the tree of life; matching profiles identifies proteins with similar patterns of inheritance. When putting all the profiles parallel to one another, each row represents a single protein while each column represents a single genome and the blackness of each element represents the degree of homology between the human gene and orthologue from that particular organism (Figure III-6). Note that no homology in either primary sequence, or in tertiary structure, is required among the proteins with similar BLASTp profiles in this approach to reveal a possible functional relationship between two proteins.

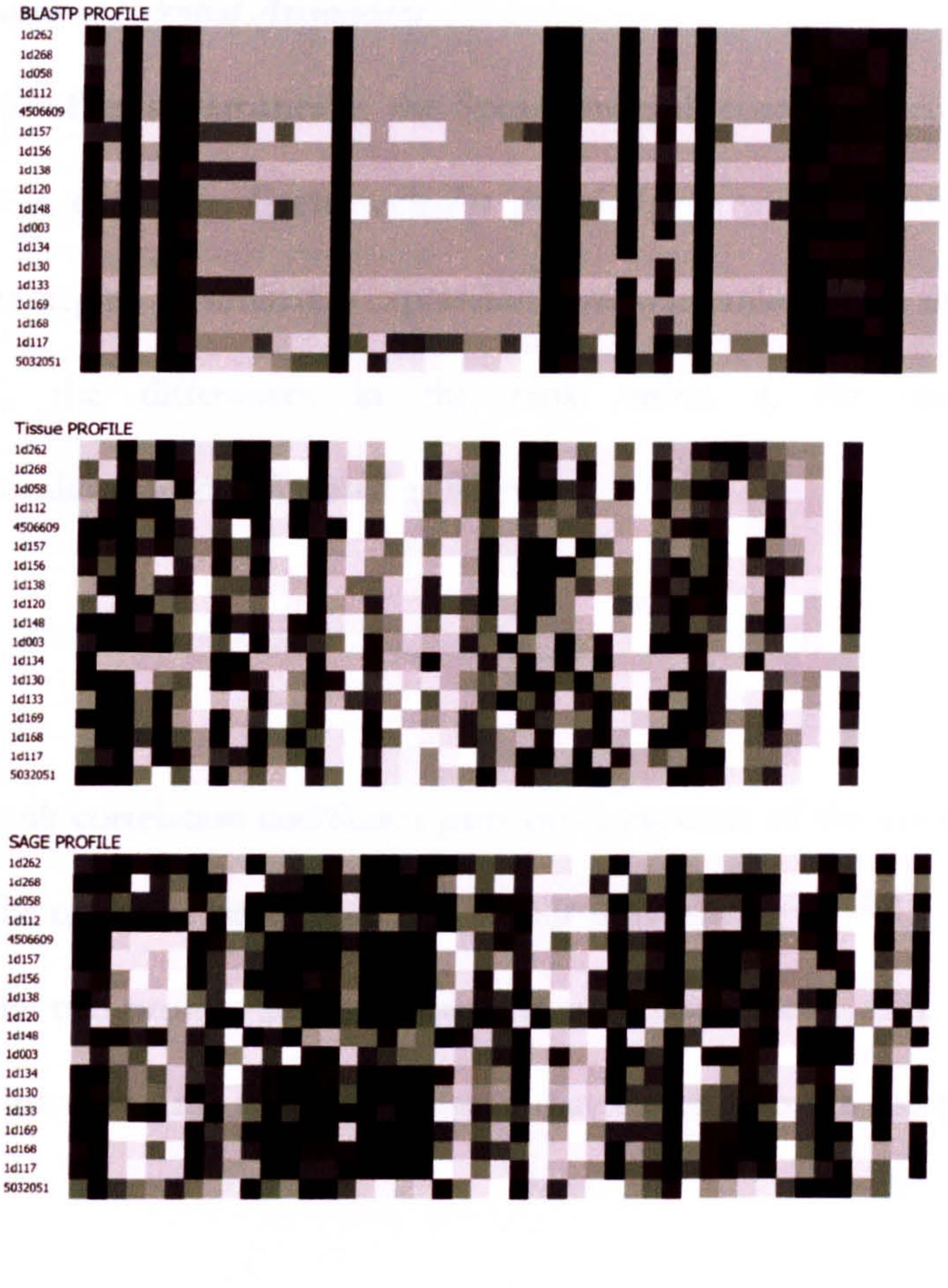


Figure III-6 **Nucleolar proteome Profiling**. The proteome profiled in terms of (top) its homology to other model organisms (BLASTp), (middle) of where they are expressed in different tissues (Unigene) and (bottom) of how they were expressed in various cells (SAGE). The data presented here represent a subset of 20 randomly chosen ribosomal proteins (See supplementary Table III-S2 for details)

Likewise, proteins are often expressed at either the same time, or place, as functionally related proteins. By either choosing different cell tissues, or by varying the growth conditions of the cells, enough variation in gene expression can be observed to identify co-expressing genes. By mining public databases that store mRNA expression levels, e.g., either expressed sequence tag (EST) libraries (Boguski and Schuler, 1995), or serial analysis of gene expression, i.e. SAGE (Lash et al., 2000), similar profile clustering was performed on each nucleolar gene for 41 different tissues and 115 different growth conditions, as illustrated in Figure III-6.

III.4.3 Visualisation of Potential Associations

To analyse the profiles systematically, the Spearman rank correlation coefficient, r_s, was calculated between each pair. Each nucleolar protein/gene was compared in a pairwise manner: first, the degree of similarity/expression level was ranked from the highest to the lowest; second, the differences in the rank order, d, for each n different species/tissue/condition were calculated as follows:

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n}$$

The Spearman rank correlation coefficient provides a measure of the association between two profiles of the nucleolar proteins/genes, with its limits being -1 and +1. If it takes the value +1, then the two profiles are the same; if it takes the value -1, then the profiles are opposite to one another. If it is zero, it implies that the profiles are independent of one another.

To collate visually all this information about potential protein-protein associations, a 3-dimensional (3D) plot was designed to visualise the data (Figure III-7). The axes were assigned for the BLASTp, tissue and SAGE profile comparisons respectively. For example, those protein-protein associations that have very similar profiles in terms of (a) how they are inherited in different genomes, (b) where they are expressed and (c) how they are expressed, would be clustered around the (1, 1, 1) corner. Conversely, unrelated protein-protein pairs would be centred at (0, 0, 0) in the 3D space (Figure III-7). Additional information content can be added by changing the shape of the object (e.g. from sphere to square), or by changing the colour or size (Figure III-7, left). For example, known motif-motif interactions such as coiled-coiled domains and existing protein-protein interaction information, from either human, or other homologues, can be added to help evaluate the potential protein-protein associations. An example of this type of data is the two recent

high-throughput analyses of protein complex composition in *Saccharomyces cerevisiae* (Bader et al., 2001; Gavin et al., 2002; Ho et al., 2002) or in http://www.bind.ca/.

400 nucleolar proteins were identified in our current analyses and hence there could potentially be 79,800 binary protein-protein associations (0.5x400x399), excluding self associations. Most of the potential associations, when put in the abovementioned 3D plot, are centred at origin (0, 0, 0) with a skewing towards the (1, 1, 1) corner (Figure III-7). This plot suggests that most of the proteins are functionally independent from one another, whilst only a small fraction of proteins are related to each other in terms of (a) how they are inherited in different genomes, (b) where they are expressed and (c) how they are expressed. For example, in Figure III-7, any protein-protein associations involving two

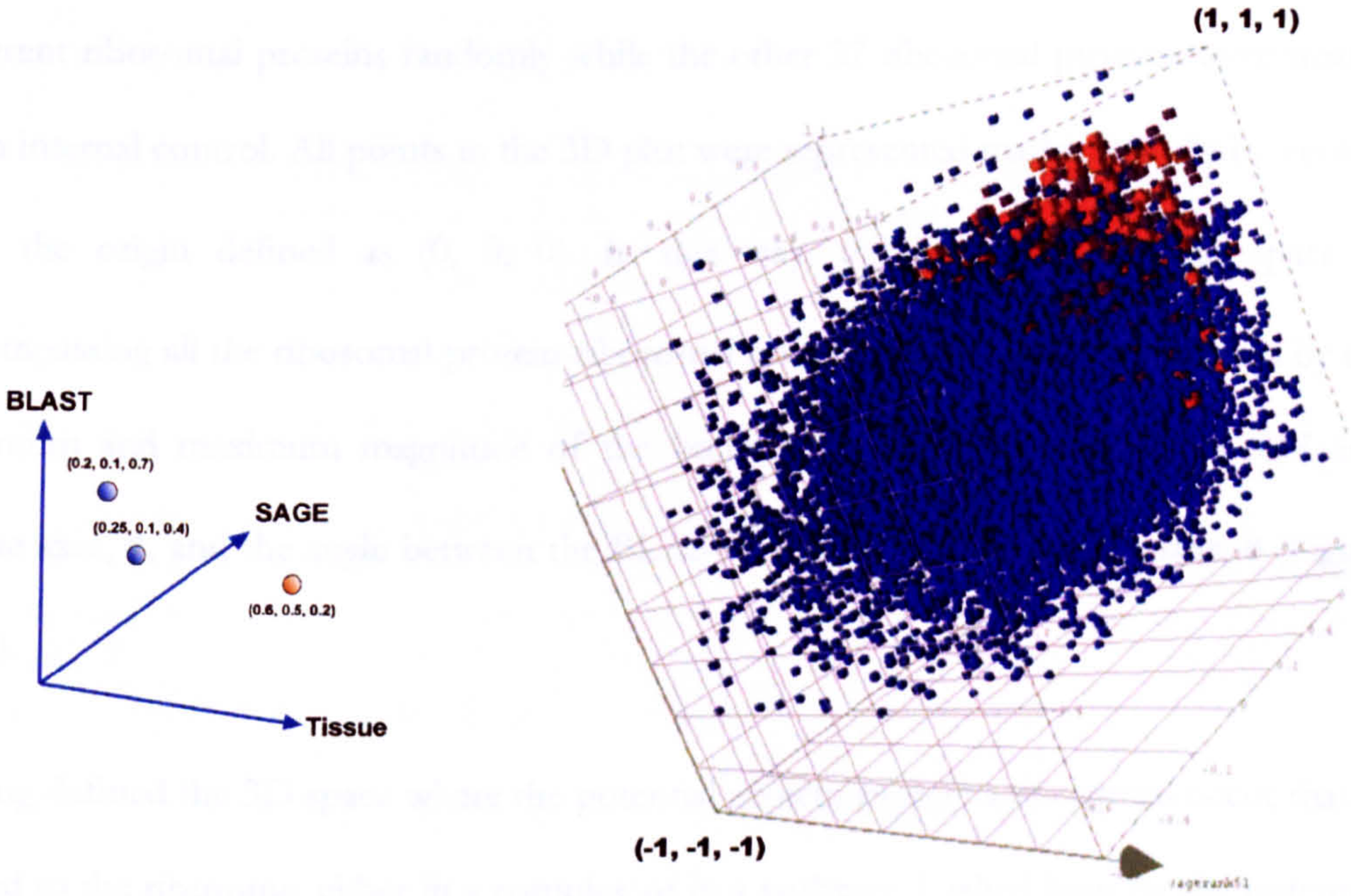


Figure III-7 **Visualisation of the complex data related to all hypothetical nucleolar protein-protein associations.** Different colour/shape can be changed for each spot to provide more dimensions of data on top of the 3D space (left). 79,800 potential protein-protein associations of the nucleolar proteome containing 400 proteins were shown. Ribosomal protein-ribosomal protein associations are highlighted in red (right).

ribosomal proteins were highlighted in red (i.e. those possible associations found in the ribosome complex) and they are clustered at the (1, 1, 1) corner.

III.5 In silico classification of Novel proteins

The clustering of known ribosomal protein-ribosomal protein associations on the (1, 1, 1) corner prompts me to investigate whether it is possible to use this clustering in classifying proteins, especially the novel proteins. The idea behind this is that protein-protein associations in the same complex/pathway performing a specified function have similar properties, for example, where all the points that represent these associations locate in a 3-D plot as in Figure III-7. For example, to identify all the potential associations that are related to the ribosome, I need to define first the 3D space Ω that was occupied by a subset of ribosomal protein-ribosomal protein associations. In this case, we chose 20 different ribosomal proteins randomly while the other 27 ribosomal proteins were treated as an internal control. All points in the 3D plot were represented mathematically by vectors with the origin defined as (0, 0, 0). In this way, the boundaries of 3D space Ω encompassing all the ribosomal protein-ribosomal protein associations were defined by the minimum and maximum magnitude of the vectors, the angle between the SAGE and Tissue axes, ϕ, and the angle between the BLASTp and the SAGE-Tissue plane, θ (Figure III-8).

Having defined the 3D space where the potential protein-protein associations occur that is related to the ribosome, either in a complex or in a pathway, I asked how far the potential associations between the remaining 380 proteins and any one of the 20 ribosomal proteins deviated from this restricted 3D space Ω? All vectors that fall outside Ω are weighed according to how much they are deviated from the 3D space. For instance, the deviations of the vectors representing potential associations between protein A and every ribosomal

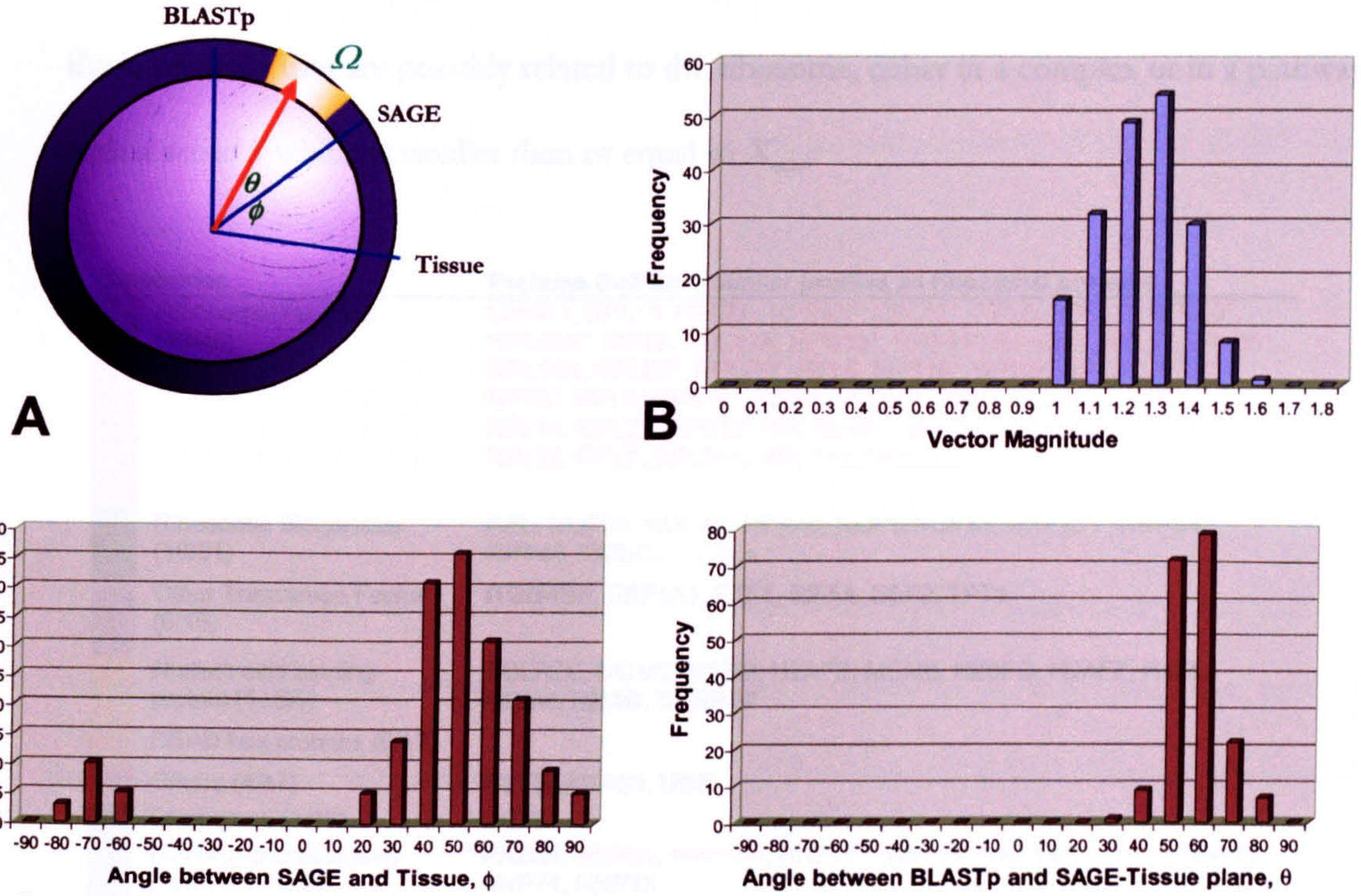


Figure III-8 **Methodologies behind the *in silico* classification.** (A) The restricted 3D space Ω defined by the subset of 21 ribosomal proteins was highlighted in orange. The distribution of the (B) magnitude and (C) the angle of those vectors encompassing the ribosomal-ribosomal protein associations within the subset.

proteins within the defined subset were computed and averaged to give a score S_A for protein A. Similarly, each ribosomal protein within the subset was compared with the other 19 ribosomal proteins and each ribosomal protein was then given a score X_R. Therefore, the range of X_R acts as an internal control to indicate how much deviations of the vectors, which represent potential protein-protein associations, inherently exist within the ribosome and this range set the *in silico* standard for other proteins to be classified as being related to the ribosome. Detailed calculation procedures can be found in the supplementary information in the accompanying CD.

The higher the score, S or X, the further away the protein-protein association is from the centre of the defined ribosomal 3D space. Hence, the maximum score derived from the set

of 20 ribosomal proteins, X_{max}, defines the upper limit of score S. Table III-5 summarises those proteins that are possibly related to the ribosome, either in a complex or in a pathway with a score S which is smaller than or equal to X_{max}.

Categories	Proteins that have similar profiles as ribosomal proteins
Ribosomal Proteins (46/46)	LAMR1, RPL15, RPL7A, RPS4X*, RPL12*, RPL5, RPL11*, RPL19*, RPL23A*, RPS8, RPL13A, RPS13*, RPS11*, RPS3A*, RPL3*, RPL26*, RPL10A, RPL32*, RPLP0*, RPL6, RPS18*, RPS9*, RPL9*, RPL4*, RPS5*, RPL8, RPS16*, RPS6, RPL18A, RPL17*, RPL35, RPL31, RPL14, RPL27, RPS12, RPL18, RPL36, RPL13, RPL34, RPL28, RPL22, RPL7, RPL35A, RPL27A, RPS27L, RPS14*
Ribosomal Biogenesis (10/31)	FBL, **NOP56**, **NOLA2**, **RPA40**, **NOP5/NOP58**, **NHP2L1**, **MRPS4**, **RRP46**, **PMSCL1**, CSL4
Other Translation Factor (6/15)	**ITGB4BP**, EEF1A1, ETF1, EIF5A, EEF2, TPT1
Nucleic acid binding protein (11/95)	POLR2E, MCM2, MCM3, H2AFE, MCM6, H2BFG, H2AFZ, H1F5, MCM4, **RNAC**, SNRPD2
DEAD box proteins (0/17)	
Others (4/47)	ACTB, ACTG1, **UBB**
Chaperone (1/16)	CCT3
Novel/uncharacterised proteins (9/123)	PA2G4, NNP89, NNP128, NNP127, **NNP10**, **NNP125**, NNP85, NNP82, NNP74, NNP15

Remarks:-

1. Ribosomal proteins or known homologues of ribosomal proteins in yeast are coloured in red and those ribosomal proteins selected to form the basis of ribosomal protein-ribosomal protein association are marked with an asterisk (*).
2. Proteins involved in ribosomal biogenesis are coloured in green;
3. Yeast or other homologues that are involved in ribosomal biogenesis are shown in **bold**;
4. The number next to each category (a/b) where a is the number of proteins that have similar profiles as ribosomal proteins and b is the total number of proteins in that category in the nucleolar proteome
5. The following 10 proteins were not included in this study due to incomplete dataset: RPS7, ASNA1, KHDRBS3, ASE-1, NNP97, SENP3, SMC4L1, NNP2, NOLA3, NNP1

Table III-5 **Summary of proteins predicted to be related to ribosomes, either within the same complex or in a pathway involving ribosomes.**

87 out of 390 proteins in the nucleolar proteome were classified as being related to ribosomes based on their BLASTp, tissue and SAGE profile comparisons. 10 proteins, including ribosomal protein RPS7, were excluded from the analysis due to incomplete datasets. All the remaining ribosomal proteins were identified in Table III-5, suggesting that this method could identify other components within the ribosome complex based on their profile-profile comparisons. On the other hand, more than one-third of the ribosomal biogenesis and translation factors were selected from the nucleolar proteome. This is reassuring because these factors are known to be related to either the formation and/or function of ribosomes.

Interestingly, none of the 17 DEAD box proteins were selected using this classification method. For some of the proteins, such as histones, replication factors and actins, their relationships with ribosomal proteins remain unclear. Unexpectedly, Ubiquitin (UBB) was classified to be related to ribosomes based on the profile-profile comparisons. Indeed, several ribosomal proteins have previously been shown to be co-transcribed with the ribosomal gene and poly-ubiquitination is important for the function for ribosomes during the cell cycle (Spence et al., 2000).

There are 9 novel proteins that were included in this *in silico* classification (Table III-6). Among those, three are clear homologues of yeast ribosomal proteins and another four proteins have potential homologues in different complexes involving in ribosomal biogenesis as illustrated in Figure III-4. For example, yeast homologues of NNP10 and NNP125 are involved in the exosome and the formation of early pre-60S complexes, respectively. NNP127 is a human homologue of yeast histone H3. NNP82, NNP74 and

Rank	Gene Name	Description	Yeast GN	Sc	Sp	Ce	Dm	Complex
44	NNP89	Novel Nucleolar Protein 89	RPL30					Ribosome
48	NNP128	Novel Nucleolar Protein 128	RPL23B					Ribosome
51	NNP127	Novel Nucleolar Protein 127	HHT2					
52	NNP10	Novel Nucleolar Protein 10	RRP4					Exosome
70	NNP125	Novel Nucleolar Protein 125	RPF2					D, F
72	NNP85	Novel Nucleolar Protein 85	RPL21A					Ribosome
75	NNP82	Novel Nucleolar Protein 82	YGR283C?					
80	NNP74	Novel Nucleolar Protein 74	SKI6?					Exosome?
84	NNP15	Novel Nucleolar Protein 15	RRP45?					Exosome?

Table III-6 **Summary of the novel proteins included in this *in silico* classification.** The lower the value of 'Rank', the less its mean distance of deviation from the 3D space Ω. 'Yeast GN' shows the closest match of a yeast gene with the corresponding proteins. The shaded panels represent the results of BLASTp searches for each protein against species-specific proteome database from EBI/SGD databases: *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Drosophila melanogaster* (Dm) and *Caenorhabditis elegans* (Ce). The results are shaded such that a black panel represents the expectancy value (e-value) of 0.0 and thus indicates a nearly perfect match; 80% grey for e-value < $1e^{-100}$; 50% for $1e^{-100}$ < e-value < $1e^{-75}$; 40% for $1e^{-75}$ < e-value < $1e^{-50}$; 25% for $1e^{-50}$ < e-value < $1e^{-25}$ and a white panel for e-value > $1e^{-25}$. 'Complex' indicates the name of the complex as illustrated in Figure III-4. It should be noted that the e-value between known yeast homologues of human proteins can be as low as $1e^{-10}$.

NNP15 do not have clear homologues in yeast, but are rather higher organism-specific. Whether they are related to the ribosome complex, or to a pathway involving the ribosome, remains to be established.

III.6 Discussion

In this chapter I reported and compiled the previously known and recently identified nucleolar proteins in human cells (c.f. Table III-1 & Table III-3). This accounts for ~ 400 proteins, which are encoded by roughly ~0.5-2.5% of the human genome, if the current estimate of 15-80,000 human genes is correct, from an organelle that has been studied for more than 150 years. Even from this small sampling from a well-characterised structure, ~30% of the proteins are still defined as either novel, or previously uncharacterised, according to their genome annotations. However, by mining the databases publicly available (Table III-2), we could get a glimpse into the functions within the nucleolus.

The nucleolus is known to be the place where ribosomal biogenesis occurs; however, the pathway of ribosome synthesis in human has not been studied in detail so far. By comparison with the known pathway from yeast (Figure III-4), mainly *Saccharomyces cerevisiae*, with its human homologues, 1/3 of the novel or uncharacterised proteins were annotated to be related to ribosomal biogenesis (c.f. Table III-4). The future analyses of these putative ribosomal biogenesis factors based on homology to their yeast counterparts, i.e. ~20% of the nucleolar proteome, should advance our understanding of how the ribosome forms in human cells.

So, what are the functions of the remaining 80% proteins identified from the nucleolar proteome? It is surprising that the addition of 136 new proteins identified from LC-MS/MS data does not change the overall distribution of proteins in different categories present in the nucleolus (c.f. Figure III-3). DNA and RNA binding proteins that are

involved in DNA repair, transcription, the unwinding of nucleic acid and RNA modification including splicing, constitute roughly one-quarter of the nucleolar proteome. The high protein complexity of the nucleolus implies that either the biogenesis of ribosomes is a surprisingly complex process and/or that the nucleolus carries out additional functions, consistent with the theory of a plurifunctional nucleolus (Pederson, 1998).

It is interesting to note the high abundance of RG-rich domains within the nucleolus and that 30% of the nucleolar proteins contain at least one RG[RG] motif (c.f. Section III.4.1.3). Previous studies have suggested that SMN, a component of Cajal bodies, utilises this domain to interact with a range of proteins involved in pre-rRNA processing, ribosome production, pre-mRNA splicing, transcription and recruitment to Cajal bodies. The close association of Cajal body with the nucleolus has been well established since its discovery (See Chapter I). It is noted that Cajal body components containing this RG-rich domains, such as Sm proteins, have been shown to be trafficking through the nucleolus (Sleeman and Lamond, 1999). It remains to be established whether such trafficking behaviour is due to the involvement with various RG-containing proteins within the nucleolus.

In general, nucleolar proteins show reduced levels of amino acids that could be potentially phosphorylated, such as serine, threonine and tyrosine, in their composition as compared with those that localised elsewhere in the nucleus (c.f. Section III.4.1.3). On the other hand, nucleolar proteins are more likely than any other nuclear proteins to have PP1-binding domains (c.f. Section III.4.1.3). In fact, from the current analyses, we identified all three isoforms of PP1 within the nucleolus (c.f. Table III-3). It will be interesting to explore in the future whether the nucleolus is a preferential site for dephosphorylation within the nucleus.

To survive, cell requires energy and nucleotides such as ATP and GTP are utilised as the major currency. For example, nuclear import requires the establishment of a GTP/GDP gradient across the nuclear envelope and DEAD box helicases require ATP to unwind DNA/RNA. In this analysis, we identified more than one ATP transporters, such as ASNA1, a human homologue of a bacterial arsenite ATP-binding transporter arsA (c.f. Table III-3). Similarly, putative GTPases such as NNP47, NNP51 and NNP57/NGB, are proposed to be involved in ribosomal biogenesis (c.f. Figure III-4 & Table III-4).

Proteins rarely operate alone, but usually function together in a network/complex. To understand more about how proteins function together in the nucleolus, I proposed a way to visualise all the potential binary protein-protein associations within the proteome. Taking advantage of the huge amount of data available from the public domain, the potential associations were also profiled according to (a) how they are inherited in different genomes, (b) where they are expressed and (c) how they are expressed (c.f. Figure III-6). The recent expanse of protein-protein interaction data in yeast, from large scale yeast-two-hybrid assays and high throughput mass spectrometry, provides a basis to evaluate the data of potential associations within the nucleolus (Gavin et al., 2002; Ho et al., 2002).

Based on the existing data, here I propose a new way to classify proteins related to the ribosome within the nucleolar proteome *in silico* (c.f. Figure III-8, Table III-5 & Table III-6). This approach can be seen as an *in silico* analogue of an "immunoprecipitation", using a set of 20 randomly chosen ribosomal proteins as bait. Within the 400 nucleolar proteins, 40% are related to ribosomes, including ribosomal proteins, proteins shown to be related to ribosomal biogenesis in human or other model organisms and translation-related factors (c.f. Figure III-3). In our *in silico* "pull down", 80% of which have been shown to be related to the ribosome, using the same criteria, we achieved therefore a two-fold enrichment over random selection. Similar analyses can be used to help classifying large

numbers of proteins in other proteomic studies to predict and reveal new protein partners either in a complex or along a pathway. However, the caveat of this approach is that it requires the related proteins to have a highly coordinated expression in different organisms, tissues and growth conditions in order to be considered as part of a complex or functioning within a pathway. This coordination may not be necessary especially if particular protein is only required in certain species or tissue to enhance functional efficiency. For example, this method weighs against those human-specific proteins that are known to be related to ribosome such as B23, nucleolin or ribonuclease P subunits. Nevertheless, this method may be able to identify the core components of a pathway/complex that are related to a specific function.

Currently, the 3D plot to visualise the potential protein-protein associations has one dimension related to the genome, while the other two relate to transcription (c.f. Figure III-7). However, the proteome is nothing but dynamic in that it changes at different stages of the cell cycle and under different metabolic conditions (c.f. Table III-1b-c). In future, it would become more powerful in classification/analyses if data concerning protein copy number under different conditions are compared and represented in a separate dimension such that three axes are independent from one another and represent the genome, transcriptome and proteome, respectively.

III.7 References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J Mol Biol. 215:403-10.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-402.

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. Curr Biol. 12:1-11.

Bader, G.D., I. Donaldson, C. Wolting, B.F. Ouellette, T. Pawson, and C.W. Hogue. 2001. BIND--The Biomolecular Interaction Network Database. Nucleic Acids Res. 29:242-5.

Bailey, T.L., and C. Elkan. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 2:28-36.

Boguski, M.S., and G.D. Schuler. 1995. ESTablishing a human transcript map. Nat Genet. 10:369-71.

Fatica, A., and D. Tollervey. 2002. Making ribosomes. Curr Opin Cell Biol. 14:313-8.

Fox, A.H., Y.W. Lam, A.K. Leung, C.E. Lyon, J. Andersen, M. Mann, and A.I. Lamond. 2002. Paraspeckles. A novel nuclear domain. Curr Biol. 12:13-25.

Gavin, A.C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 415:141-7.

Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, and M. Tyers. 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 415:180-3.

Lash, A.E., C.M. Tolstoshev, L. Wagner, G.D. Schuler, R.L. Strausberg, G.J. Riggins, and S.F. Altschul. 2000. SAGEmap: a public gene expression resource. Genome Res. 10:1051-60.

Marchler-Bauer, A., A.R. Panchenko, B.A. Shoemaker, P.A. Thiessen, L.Y. Geer, and S.H. Bryant. 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 30:281-3.

Paushkin, S., A.K. Gubitz, S. Massenet, and G. Dreyfuss. 2002. The SMN complex, an assemblyosome of ribonucleoproteins. Curr Opin Cell Biol. 14:305-12.

Pederson, T. 1998. The plurifunctional nucleolus. Nucleic Acids Res. 26:3871-6.

Pellegrini, M., E.M. Marcotte, M.J. Thompson, D. Eisenberg, and T.O. Yeates. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A. 96:4285-8.

Scherl, A., Y. Coute, C. Deon, A. Calle, K. Kindbeiter, J.C. Sanchez, A. Greco, D. Hochstrasser, and J.J. Diaz. 2002. Functional proteomic analysis of human nucleolus. Mol Biol Cell. 13:4100-9.

Sleeman, J.E., and A.I. Lamond. 1999. Nuclear organization of pre-mRNA splicing factors. Curr Opin Cell Biol. 11:372-7.

Spence, J., R.R. Gali, G. Dittmar, F. Sherman, M. Karin, and D. Finley. 2000. Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. Cell. 102:67-76.

Terns, M.P., and R.M. Terns. 2001. Macromolecular complexes: SMN--the master assembler. Curr Biol. 11:R862-4.

CHAPTER IV

DYNAMICS OF SUBNUCLEOLAR DOMAINS

*"In the beginning God created the heavens and the earth. The earth was without form, and void; and darkness was on the face of the deep. And the Spirit of God was hovering over the face of the waters. Then God said, "**Let there be light**"; and there was light. And God saw the light, that it was good; and God divided the light from the darkness"*

~ Genesis, The Bible ~

IV. Dynamics of Subnucleolar Domains

IV.1 Introduction

As discussed in the last chapter, the protein composition within the nucleolus is very complex and the nucleolus is composed of at least 400 members. Yet, the nucleolus is not a homogenous soup of proteins; instead it has at least 3 different subnucleolar domains for defined functions. The subnucleolar domains, namely the fibrillar centre (FC), the dense fibrillar component (DFC) and the granular component (GC) (See Chapter I), were originally defined by their morphology and differential staining in the Transmission Electron Microscope (TEM). These subcompartments are interlinked: multiple FCs are surrounded by the DFC while the GC is the outermost region engulfing the FC/DFC regions. Subsequent immunostaining studies, using different antibodies raised against nucleolar antigens, have helped to define the functions of these morphologically distinct subnucleolar domains. For example, subunits of RNA polymerase I, that are responsible for rDNA transcription, and its related transcription initiation and termination factors are present in FCs, whilst rRNA processing factors such as fibrillarin, NHPX and GAR1 are localised in DFCs and ribosomal proteins and chaperones are found in the GC (reviewed in Shaw and Jordan, 1995 and see below Figure IV-1). It was then hypothesised that ribosomal DNA transcription took place in FCs (or at the FC/DFC boundary) and that rRNA transcripts were subsequently processed and modified in the DFC (Huang, 2002; see Chapter I). The processed rRNA transcripts were then assembled with ribosomal proteins in the GC to form ribosomal subunits (Shaw and Jordan, 1995). The proposed functions of the respective subnucleolar domains are summarised in Table IV-1.

During interphase nucleoli are stable structural entities which can be isolated in pure form for functional studies. However, they are disassembled in the transition from G2 to M phase and later precisely re-formed into the defined pattern of subnucleolar domains after

Subnucleolar Domains	Functions	Components tagged
Fibrillar Centre (FC)	Ribosomal DNA transcription	RPA39 (C/Y/G/G3)
Dense Fibrillar Component (DFC)	rRNA processing including 2'-O-methylation and pseudouridylation	FIB (C/Y); NHPX (C/Y)
Granular Component (GC)	Ribosomal proteins assemble with processed rRNA transcripts with the assistance of chaperones.	B23 (C/Y/D*) RL27 (C/Y/D*)

Table IV-1 **Summary of known functions in 3 distinct subnucleolar domains.** Column 'Components tagged' indicates the proteins that were tagged with fluorescent proteins in this study; C for cyan fluorescent protein, Y for yellow fluorescent protein, G for green fluorescent protein, G3 for tandem triple green fluorescent proteins, D for red fluorescent protein DsRed2 and * indicates the tagged versions that I have not been able to isolate as stable cell lines.

every cell division in most of the eukaryotes. However, how the subnucleolar domains break down and re-form in relation to one another remains unknown. In this chapter, I report the use of a battery of HeLa cell lines stably expressing one or more nucleolar factors labelling distinct domains to examine (1) the timing of how newly transcribed ribosomal RNAs move through nucleoli in living cells and (2) how each subnucleolar structural domain disassembles and re-forms dynamically during mitosis. I propose a model that explains how the size and the number of nucleoli are defined by the interplay between the differential localisation of nucleolar factors and chromatin.

IV.2 Generation of HeLa cell lines expressing FP-tagged nucleolar factor(s)

IV.2.1 HeLa cell lines stably expressing a single FP-tagged nucleolar factor

The original definitions of different subnucleolar domains are based on the staining in fixed EM samples. In order to study their dynamic properties in living cells, I tagged 5 different nucleolar proteins as markers for distinct subnucleolar domains, in some cases, with 4 different fluorescent protein tags, i.e. Cyan, Yellow, Green and Red (DsRed2) fluorescent proteins (Table IV-1 and data not shown). 12 different stable HeLa cell lines, labelling individual subnucleolar domains with different fluorophores, were thereby generated and these are summarised in Table IV-1. The details of the cloning and the cell

line generation were documented in Chapter II and an example of 3 cell lines labelling distinct subnucleolar domains is illustrated in Figure IV-1. The generated cell lines were subsequently characterised with respect to their fluorescent patterns, their cell cycle behaviour, their expression levels, their localisation patterns in TEM and, if possible, also at the functional level.

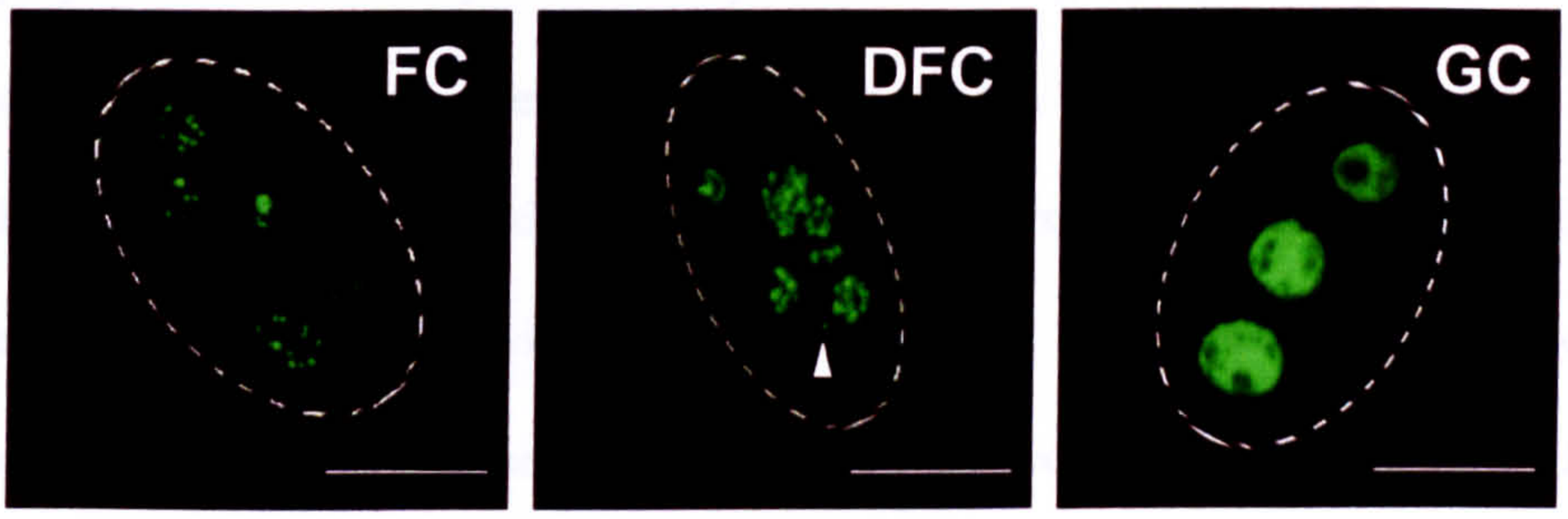


Figure IV-1 **Labelling of the 3 distinct subnucleolar domains in living cells**. The proteins tagged for FC, DFC and GC shown here are RPA39, NHPX and B23, respectively. The dotted ovals define the boundaries of nuclei. Most of the DFC components are localised both in nucleoli and Cajal bodies (white arrowhead). Some of the GC proteins such as PWP1 and ribosomal protein RL27 are localised both in nucleoli and the cytoplasm. Scale bar = 5 μm.

The fluorescent pattern of each subnucleolar domain was verified with the corresponding antibody staining and, in each case, all showed identical pattern (data not shown; e.g. see below Figure IV-7 and Chapter V). In the case of ribosomal protein RL27 and RNA polymerase I subunit RPA39, because the cognate antibodies were not available, antibodies raised against ribosomal protein RPS6 and RPA20 were used instead, respectively. Subsequent TEM studies (the work of Dr. C Lyon) using anti-GFP antibodies demonstrated that the FP-tagged ribosomal protein RL27 was found in nucleoli, ER membranes and in the cytoplasm, while the FP-tagged polymerase I subunit RPA39 was found specifically in fibrillar centres (Figure IV-2A and data not shown).

To verify that the FP-tagged polymerase I subunit is functional, immunocomplexes isolated using anti-GFP antibodies from nuclear extracts of the FP-tagged RPA39 cell line were tested *in vitro* for RNA polymerase I activity (Figure IV-2B). Indeed, the

immunocomplexes isolated from nuclear extracts of the FP-tagged RPA39 cell line supported *in vitro* RNA polymerase I transcription (Figure IV-2B, lanes 3-7) whilst negative controls omitting the anti-GFP antibodies during the immunoprecipitation (lane 1), or using nuclear extracts of parental HeLa cell (lane 2) did not (The work of Dr. G Miller). Because the transcription factors UBF and SL1 were known to be loosely associated with RNA polymerase I holoenzyme and may be lost during the immunoprecipitation, the addition of recombinant proteins increased the RNA polymerase I activity, further strengthening the conclusion that the immunocomplexes specifically contained RNA polymerase I transcription activity. (Figure IV-2B, lanes 4-7). It was noted that extracts from both the parental HeLa and the FP-tagged RPA39 cell line had similar level of *in vitro* RNA polymerase I transcriptional activity (lanes 8 and 9, respectively). Therefore, because both nuclear extracts were active for transcription, while only the immunocomplexes from the extract of the FP-tagged RPA39 cell line supported RNA polymerase I-specific transcription, I conclude that the FP-tagged polymerase I subunit RPA39 must be functional.

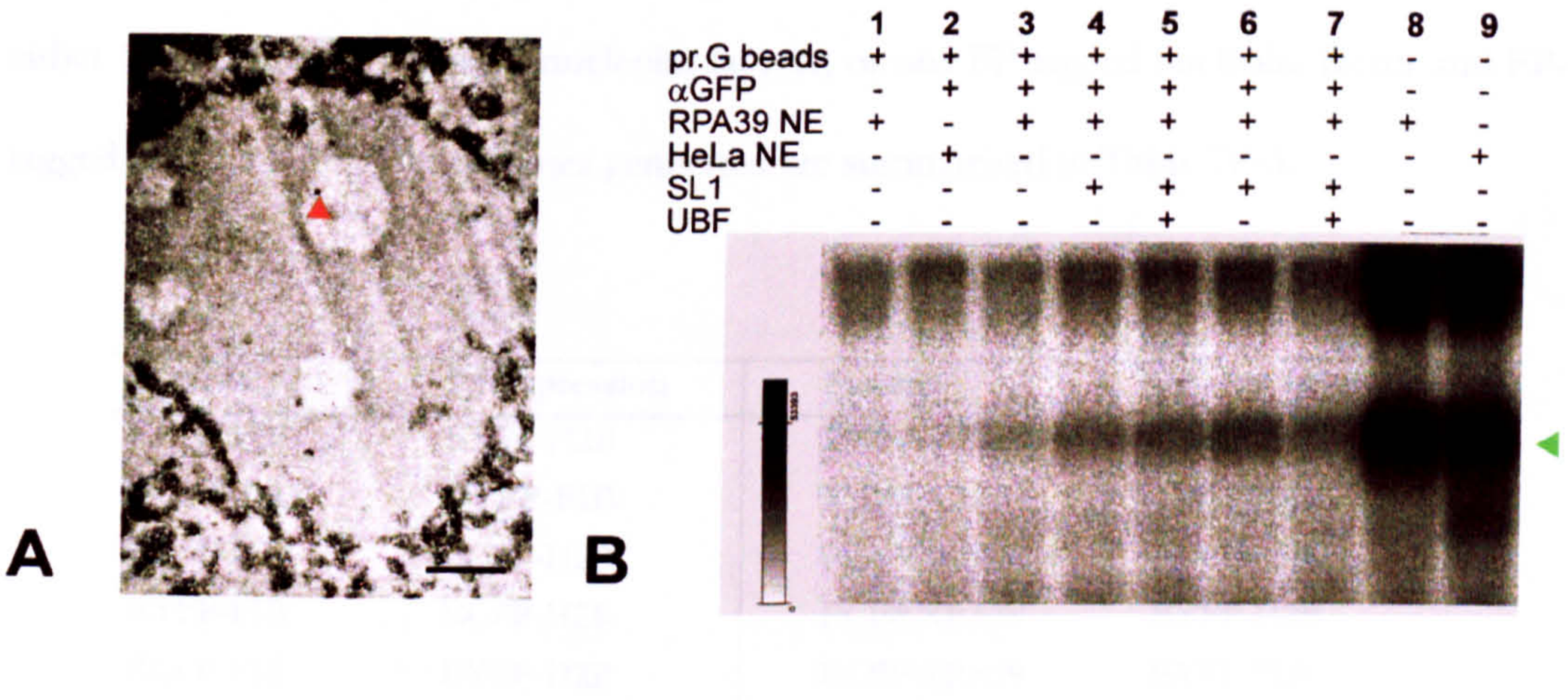


Figure IV-2 **Characterisation of FP-RPA39 cell lines.** (A) TEM samples were immunolabelled with gold conjugated anti-GFP antibodies and the red arrowhead indicates that FP-RPA39 was localised in the FC. Scale bar = 2μm (B) *In vitro* RNA polymerase I assay. The green arrowhead indicates the presence of specific transcription product. For assays involving immunoprecipitation, anti-GFP antibodies (αGFP) were coupled to protein-G beads (pr. G beads) and used for immunoprecipitations from Hela nuclear extracts (HeLa NE) or the nuclear extract made from HeLa cells stably expressing FP-RPA39 (RPA39 NE). The immunoprecipitates were washed with PBS thrice prior to performing the *in vitro* assays (See Miller et al., 2001 for details).

Examining the DNA content of the cells by FACS analyses had shown that, for all cell lines tested, the FP-tagged proteins did not block cell cycle progression. The characterisations of the cell lines are summarised in Table IV-2. A detailed example of cell line characterisation is also illustrated for FP-tagged NHPX in Chapter V.

Cell line	Immunostaining	FACS	Expression	TEM
RPA39	(√)	√	(√)	√
FIB	√	√	ND	√
NHPX	√	√	√	ND
B23	√	√	√	√
RL27	(√)	ND	-	√

Table IV-2 **Characterisation of HeLa cell lines expressing FP-tagged nucleolar proteins.** *Related antibodies* were used for immunofluorescence studies if the antibody raised against a particular antigen was not available as detailed in text and indicated here as (√). The expression of fluorescent protein tagged RPA39 was examined with a RNA polymerase I transcription assay as detailed in text. A tick under a column indicates the assay has been performed for the cell line indicated on the left. 'ND' indicates the result for the assay on that column was not determined. '-' indicates that the experiment is not possible due to the unavailability of the reagents.

IV.2.2 Generation of HeLa cell lines expressing more than one FP-tagged marker

To observe more precisely how distinct subnucleolar domains move in live cells, both in relation to one another and to chromatin, I generated 12 HeLa cell lines stably expressing either two different FP-tagged nucleolar factors, or one FP-tagged nucleolar factor and FP-tagged histone H2B. The cell lines generated are summarised in Table IV-3.

Parental	2nd expression	Parental	2nd expression
ECFP-B23	EYFP-H2B	**EYFP-RPA39**	**ECFP-FIB**
ECFP-B23	**EYFP-FIB**	EYFP-RPA39	ECFP-H2B
EYFP-B23	ECFP-H2B	EGFP-RPA39	EYFP-H2B
EYFP-FIB	ECFP-H2B	ECFP-RPA39	EYFP-H2B
ECFP-FIB	EYFP-H2B	ECFP-RPA39	EYFP-FIB
ECFP-FIB	EYFP-B23	ECFP-RPA39	EYFP-B23

Table IV-3 **Stable cell lines with double constructs generated in this study.** 'Parental' HeLa cell line expressing a single construct was transfected with a '2nd expression' vector. See text for details. The analyses of the pairs shown in bold have been described in this chapter.

Briefly, a 'parental' HeLa cell line grown in G418-containing medium and expressing a single FP-tagged nucleolar factor was transfected with a second construct as described in Table IV-3 using effectene (QIAGEN; Chapter II). After 48 hours of transfection, transfected cells were selected by treatment with the additional antibiotic blasticidin and subcloned as described in Chapter II for generating a HeLa cell line expressing a single construct. The strategy for generating a HeLa cell line expressing two constructs (or hereafter 'double stable' for simplicity) is summarised in Figure IV-3. Although the full characterisation of every cell line has not been performed yet, we are confident that the tagged proteins behave as their endogenous counterparts, based on our previous characterisation on the HeLa cell lines stably expressing the corresponding single construct alone.

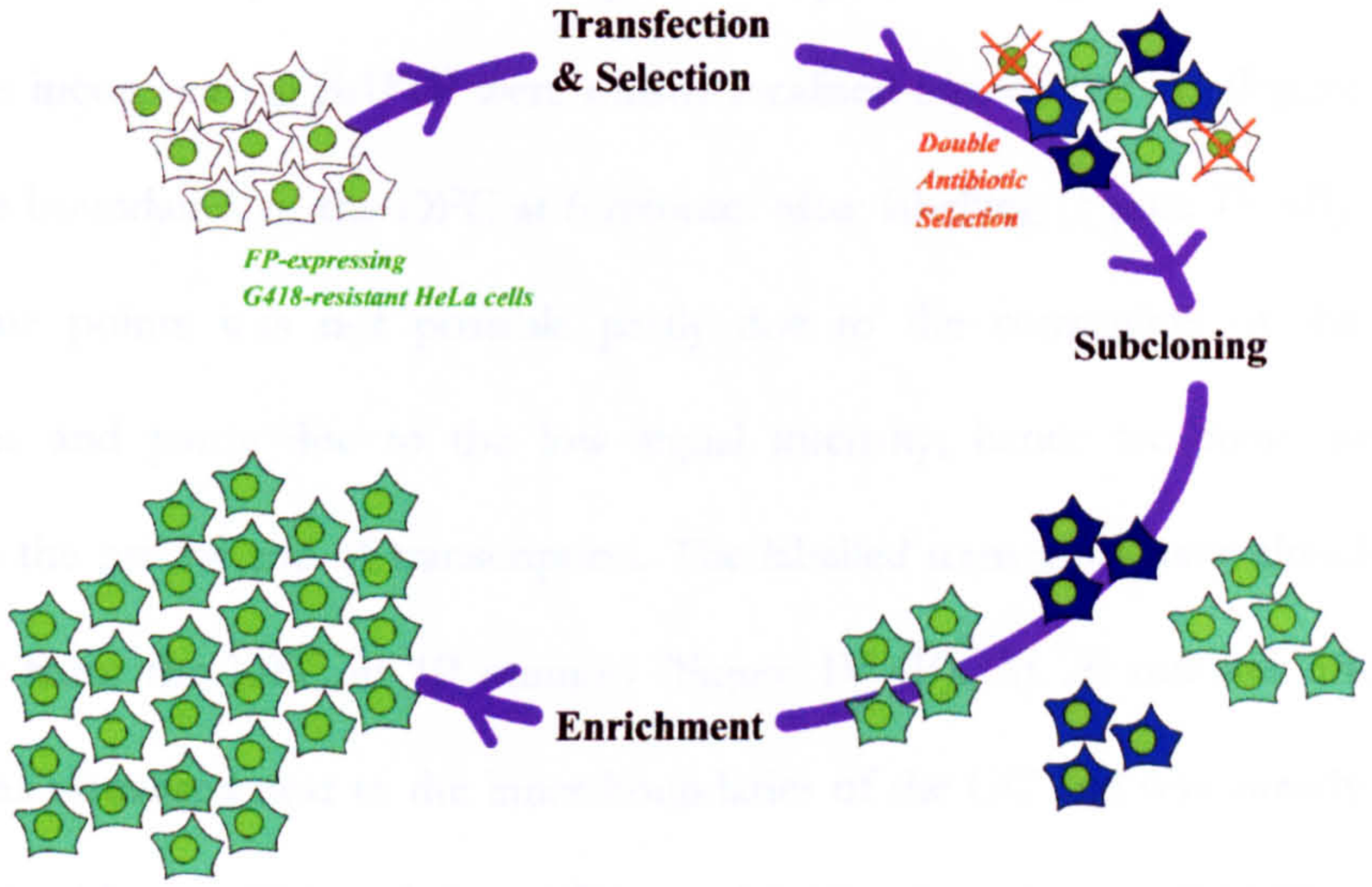


Figure IV-3 **Scheme of generating a 'double stable'.** The FP-expressing, G418-resistant HeLa cells were transfected with a construct expressing either another FP-tagged nucleolar factor or FP-tagged histone H2B. After 48 hours of transfection, the cells were selected with blasticidin and G418. Individual double antibiotic-resistant colonies were selected and picked manually under the fluorescence microscope to ensure that both FP-tagged proteins were localised properly. Cell lines with suitable fluorescent intensities were used for live cell studies.

IV.3 Timing of vectorial transport of rRNA transcripts

To understand how rRNA transcripts move from their transcription sites, I used a method of BrUTP incorporation that was developed by Dr. YW Lam in our laboratory to label the transcripts in living cells (Figure IV-4). This method differs from those previously described for permeabilised cells in that it allows BrUTP to be delivered across the cell membrane by a mild hypotonic treatment while maintaining cell viability (see legend of Figure IV-4 for the method details).

BrUTP incorporation experiments were performed in parallel on 3 stable cell lines, expressing markers for the FC (EYFP-RPA39), the DFC (EYFP-FIB) and the GC (EYFP-RL27), respectively. This is to ensure that the experiments were performed under the same growth and fixation conditions and the images collected can therefore be correlated with one another with respect to the timing of incorporation (Figure IV-4). The ribosomal transcripts incorporating BrUTP were mostly localised outside the FC (Figure IV-4A) but within the boundaries of the DFC at 6 minutes after labelling (Figure IV-4B). Analysis of earlier time points was not possible partly due to the constraints of the incubation procedures and partly due to the low signal intensity; hence we could not currently determine the precise site of transcription. The labelled transcripts have already started to move out from the DFC at 10 minutes (Figure IV-4B). At 20 minutes after labelling, BrUTP was visualised near to the inner boundaries of the GC and was already completely colocalised with the GC at 1 hour (Figure IV-4C). Cytoplasmic BrUTP staining was observed at a later time point (3 hours; data not shown) and this suggests that the BrUTP incorporated transcripts are, by this time point, likely to be processed and assembled into ribosomes. From these time-course experiments, it appears that the ribosomal transcripts move vectorially from their transcription sites (FC) towards the processing sites (DFC) and then assembly sites (GC) in an orderly fashion *in vivo*.

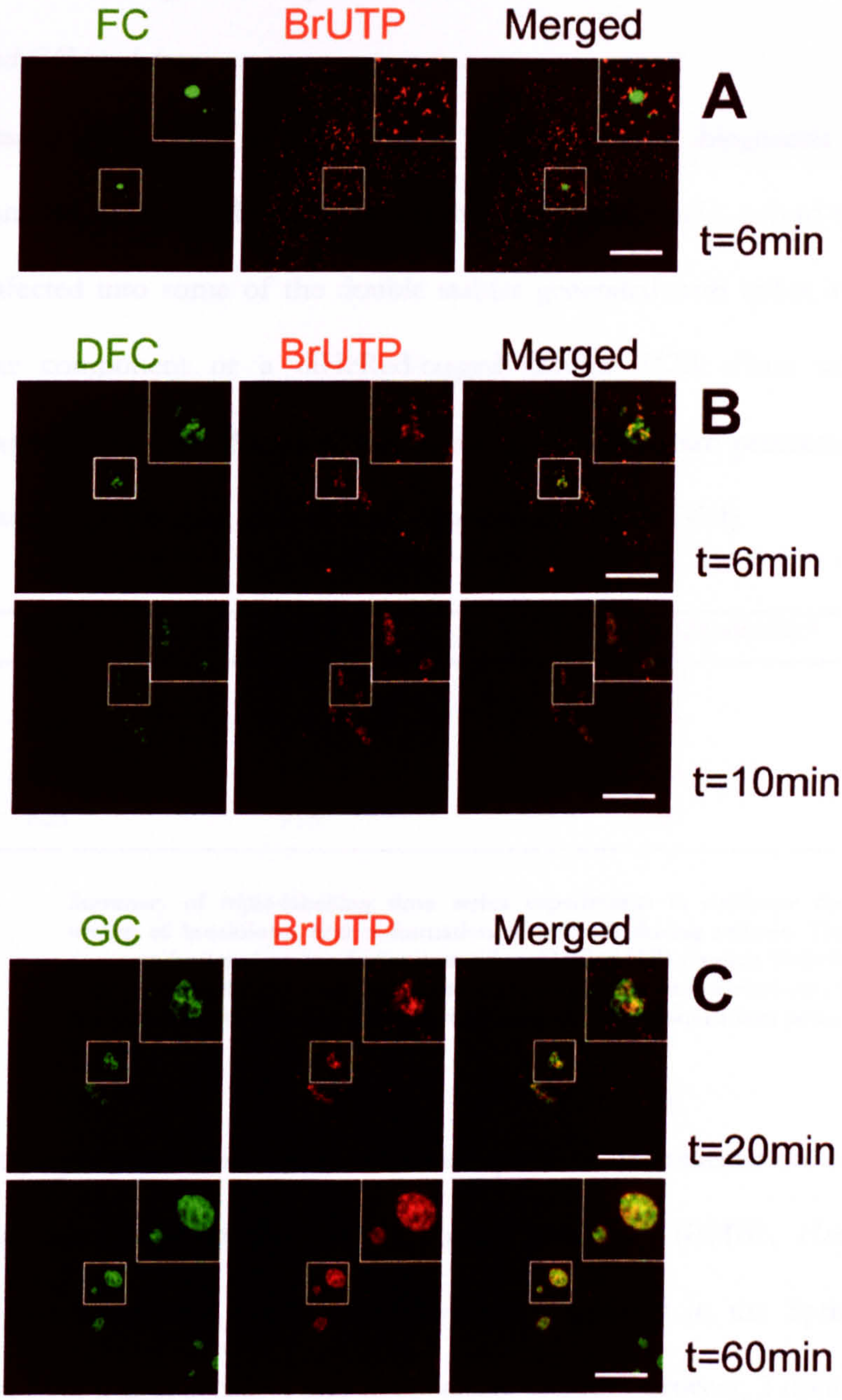


Figure IV-4 **Vectorial movement of rRNA transcript**. The subnucleolar domains (A) FC, (B) DFC and (C) GC were labelled by EYFP-RPA39, EYFP-FIB and EYFP-RL27 respectively. The middle panel illustrates the localisation of accumulated BrUTP at the defined times and the patterns of incorporation within nucleoli are shown in detail in the corresponding insets. Coverslips seeded with HeLa cells were rinsed with hypotonic KH buffer (30mM KCl, 10mM HEPES pH7.4) briefly and incubated with 50µl KH buffer containing 10mM BrUTP (Sigma) for 5 minutes in a 5% CO_2 incubator at 37°C. The cells on coverslips were 'chased' with DMEM with 20% FCS and G418 for a defined time (0, 6, 10, 15, 20, 60 and 180 minutes) in the incubator to chase the transcripts before fixation. Prior to methanol fixation for 20 minutes at -20°C, the coverslips were rinsed with PBS. The cells were then permeabilised with acetone for 30 seconds and air-dried for 10 minutes, followed by rehydration with PBS for 5 minutes and immunostaining using the anti-BrdU (1:5) antibody. The conditions developed for this method preferentially label the ribosomal transcripts, with a low visible level of RNA Polymerase II transcripts especially at shorter time points.

IV.4 On breakdown and biogenesis during mitosis

IV.4.1 DFC and GC breakdown

In order to dissect the timing of the dynamic breakdown and biogenesis of the 3 subnucleolar domains, in relation both to one another and to chromatin, a third vector was transiently transfected into some of the double stables generated with either a DsRed2-tagged nucleolar component or a diHcRed-tagged histone H2B. Four sets of '7-Dimensional' experiments (3D+Time+2 extra wavelengths+DIC) are presented here to delineate the relationship between each subnucleolar domain (Table IV-4).

ECFP	EYFP	RFP (diHcRed/DsRed2)*
FIB	RPA39	RL27
FIB	RPA39	H2B
B23	FIB	H2B
B23	FIB	RL27

Table IV-4 **Summary of triple-labelling time series experiments to delineate the timing of breakdown and re-formation of nucleoli during mitosis.** The construct for transient transfection was either diHcRed-H2B (tandem HcRed-H2B) or DsRed2-RL27. Only 2 particular double stable pairs were selected out of the 12 as these two provided the highest intensity with the minimum laser power required.

The 7D experiments presented here, unless otherwise specified, were carried out on a Zeiss LSM510 confocal microscope in Dr. Jan Ellenberg's laboratory (EMBL, Heidelberg), supported through a short-term EurALMF fellowship I received in the Spring 2002. Because mitosis is both a light-sensitive and temperature-sensitive process, 2 double stable cell lines of sufficient brightness were chosen to minimise the power of laser illuminated while maintaining a high signal-to-noise ratio. The cells were grown in a CO_2-independent imaging medium containing 0.5mg/ml L-ascorbic acid and 20%FCS for optimal growth and the imaging was performed using a temperature-regulated stage at 37°C.

When a cell is ready to enter M-phase, the first visible indicator is that the chromosomes become progressively condensed (prophase). Due to the requirement for a chromosomal

marker (diHcRed-H2B) to distinguish early prophase cells, it was not possible to compare all 3 subnucleolar domains in the same cell. This is because there are only 3 spectra-distinguishable live cell fluorophores currently available and thus two sets of 7D experiments were performed to study two subnucleolar domains each time. Figure IV-5 and Figure IV-6 show two examples of a triple labelling 3D time series studying the timing of the nucleolar breakdown and reassembly during mitosis. In both cases, the double stables were transiently transfected with diHcRed-H2B for 24 hours before imaging.

It is apparent that each subnucleolar compartment remained intact while the chromosomes condensed (Figure IV-5, 00:00-14:30 and Figure IV-6, 00:00-06:04). Possibly, when FCs disappear (Figure IV-6, 06:04-09:26), both DFCs and GCs disappeared together within a short period of time around 10 minutes (Figure IV-5, 22:54-31:48 and Figure IV-6, 09:26-19:54). I noted that immediately after the disappearance of FCs, represented by FP-tagged RNA polymerase I subunit RPA39, the DFC lost their characteristic punctate pattern (Figure IV-6, 06:04-09:26; c.f. Figure IV-1). Whether such immediate nucleolar breakdown is due to either the degradation of the factors, or their diffusions to the extranucleolar pool, remains to be determined.

In contrast with breakdown, the re-formation of both the DFC and GC compartments proceeded at different rates. DFC re-formed at around 5 minutes after anaphase, when the condensed chromosomes segregated to opposite poles (Figure IV-5, 50:52-60:53), whilst the GC took at least 20 minutes more before it reappeared (Figure IV-5, 50:52-80:40). Both timings are also different from the relocation of the RNA polymerase I subunit RPA39 after mitosis, as illustrated in Figure IV-7.

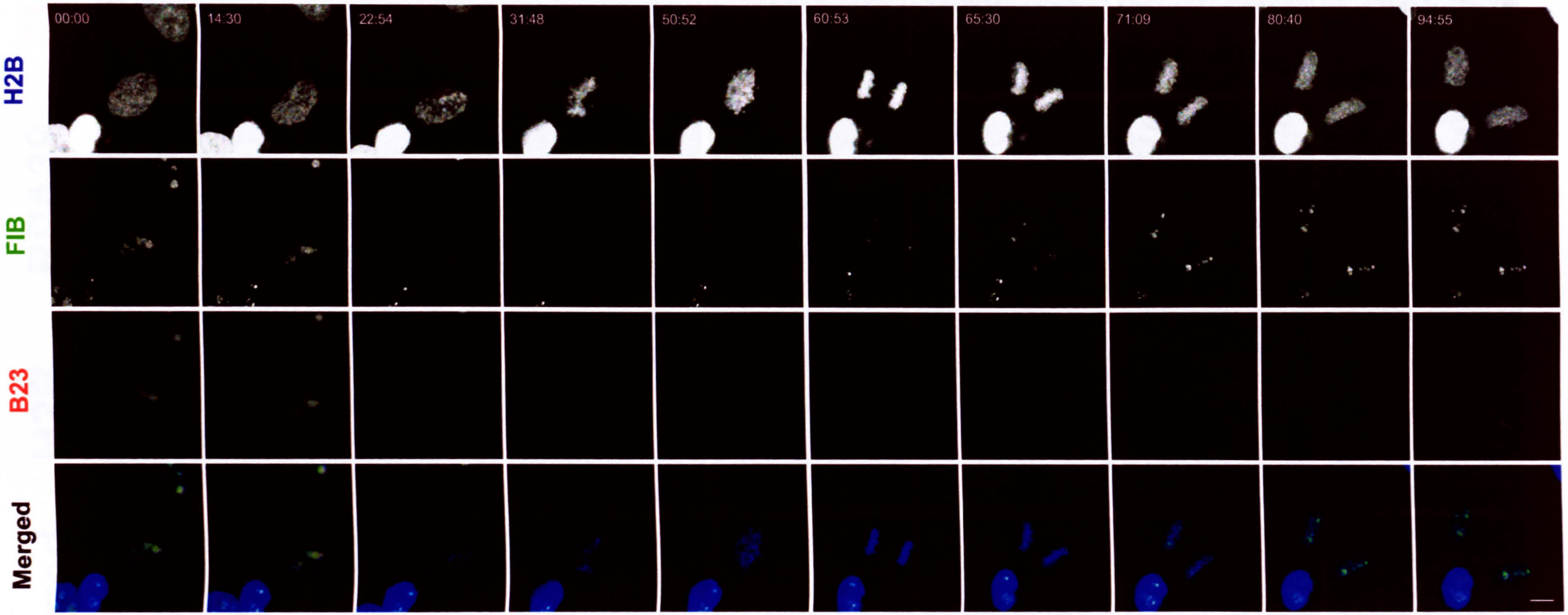


Figure IV-5 **Timing of DFC and GC breakdown and assembly.** Double stable expressing ECFP-B23 and EYFP-FIB was transiently transfected with diHcRed-H2B. Early prophase cells were identified under the eyepiece by the condensation of chromatin and followed until 40 minutes after cell division. Shown here is part of the time series of the maximal projection images of each channel at a defined time point. The whole series can be found in movie IV-5 in the supplementary CD. Time indicated = min:sec. Scale bar=5μm.

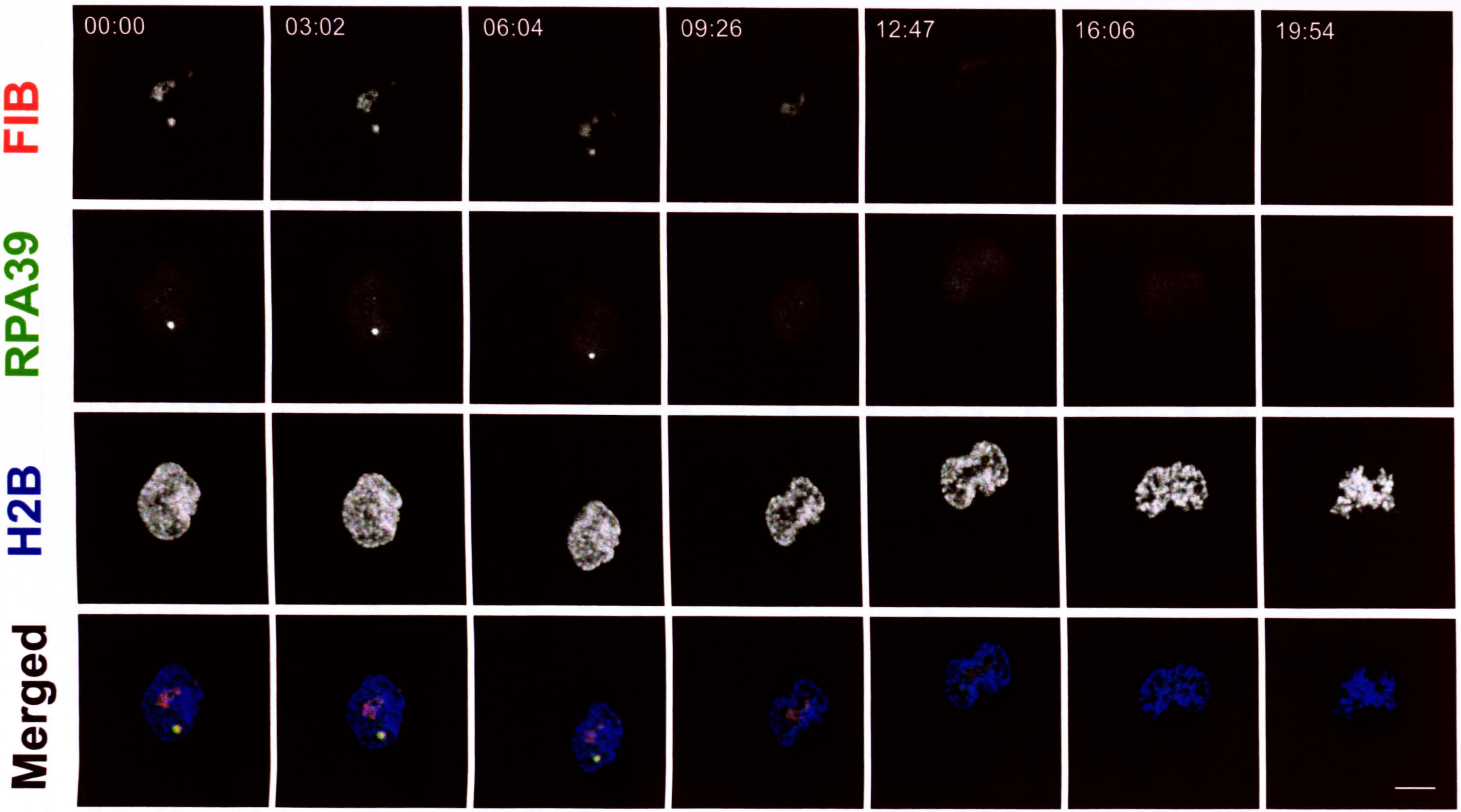


Figure IV-6 **Timing of FC and DFC breakdown.** Double stable HeLa cells expressing ECFP-FIB and EYFP-RPA39 were transiently transfected with diHcRed-H2B. Part of the time series of the projection images of each channel are shown here at a defined time point and the whole series can be found in movie IV-6 in the supplementary CD. Time indicated = min:sec. Scale bar=5μm.

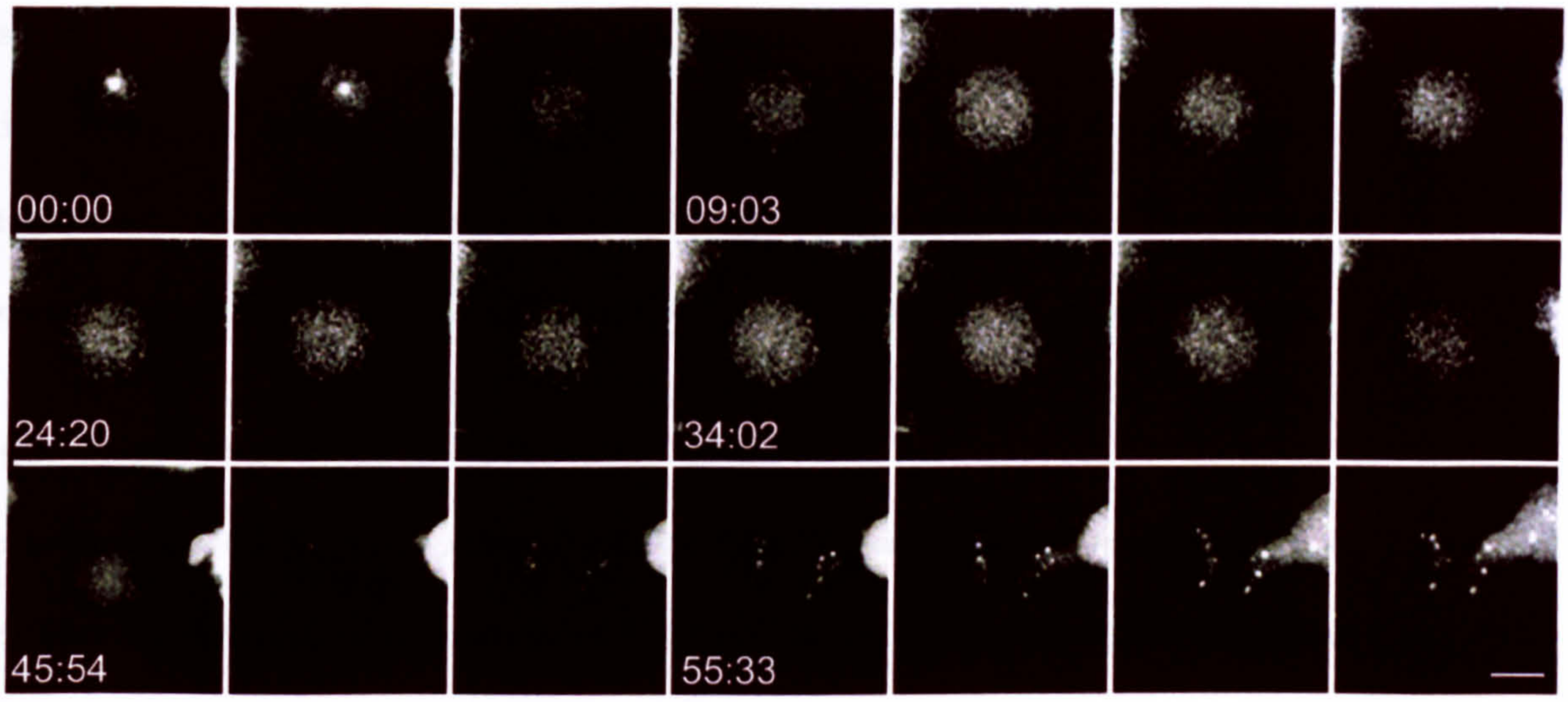
IV.4.2 Structural and dynamic FC components

Surprisingly, the polymerase I subunit RPA39 did not remain associated with chromosomes at all stages during mitosis, as previously reported (Figure IV-7A, 00:00-42:30; c.f. Scheer and Rose, 1984; Gilbert et al., 1995). Instead, the original spots diffused to the cellular pool during mitosis and did not re-form to spot-like structures (FC) until late metaphase (white arrow, Figure IV-7B). The number of FCs increased dramatically towards the end of telophase (Figure IV-7A, lowest panel). The dynamics during mitosis remained the same even when the same subunit was tagged with tandem triple EGFPs at the N-terminus to increase the resolution (data not shown). The same results were obtained both using a laser-scanning confocal microscope (Zeiss; data not shown) and a widefield epifluorescence deconvolution microscope (DeltaVision; Figure IV-7).

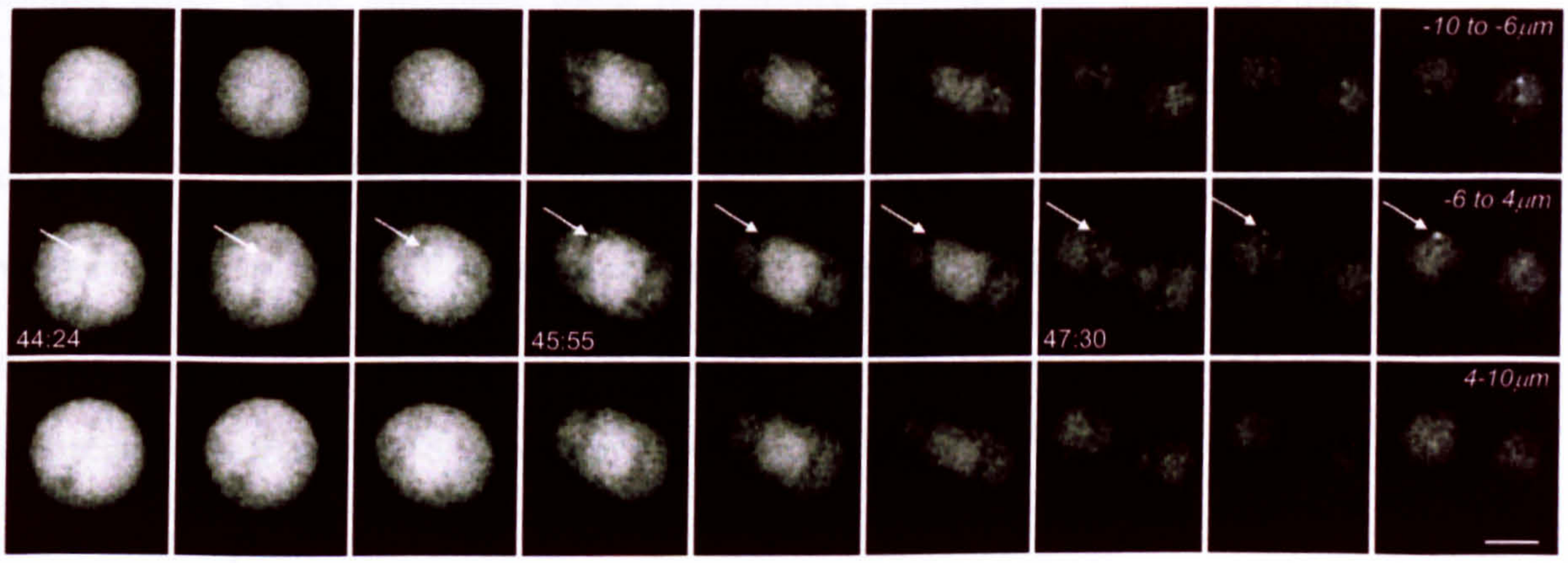
Therefore, I next investigate whether other FC proteins also behaved in this manner. I examined the localisation of the transcription factor UBF and compared this with the localisation of EYFP-RPA39 and another endogenous RNA polymerase I subunit using a monoclonal antibody, B6-2 (Figure IV-7C). During interphase, EYFP-RPA39, the endogenous RNA polymerase I and UBF colocalised in FCs (Figure IV-7C, upper). However, only UBF was observed to remain in spot-like structures (possibly the NORs) during metaphase, but neither the FP-RPA39 nor the other endogenous polymerase subunit showed this pattern (Figure IV-7C, lower). This suggests that although both UBF and RNA polymerase I subunits localised in the FC during interphase, they may have different dynamic properties during mitosis.

Figure IV-7 **A core subunit of polymerase I does not associate with chromosomes during mitosis.** (A) A time series of maximal projection imaging of the EYFP-RPA39 cell line from early prophase until G1. The images shown are maximal projections of the whole Z-series and they are taken by a DeltaVision wide-field Deconvolution microscope; every time frame = 3 minutes. (B) The same series between early anaphase and telophase are shown in detail; every time frame = 30 seconds and each image is a maximal projection between a defined depth as indicated on the right. White arrows followed the formation of a newly re-formed spot-like structure containing RNA polymerase I subunit. (C) Comparison between UBF and RNA polymerase I subunit using EYFP-RPA39 (FP-RPA39) and monoclonal antibody B6-2 raised against RPA20 during interphase (upper) and metaphase (lower). Scale bar = 5 μm.

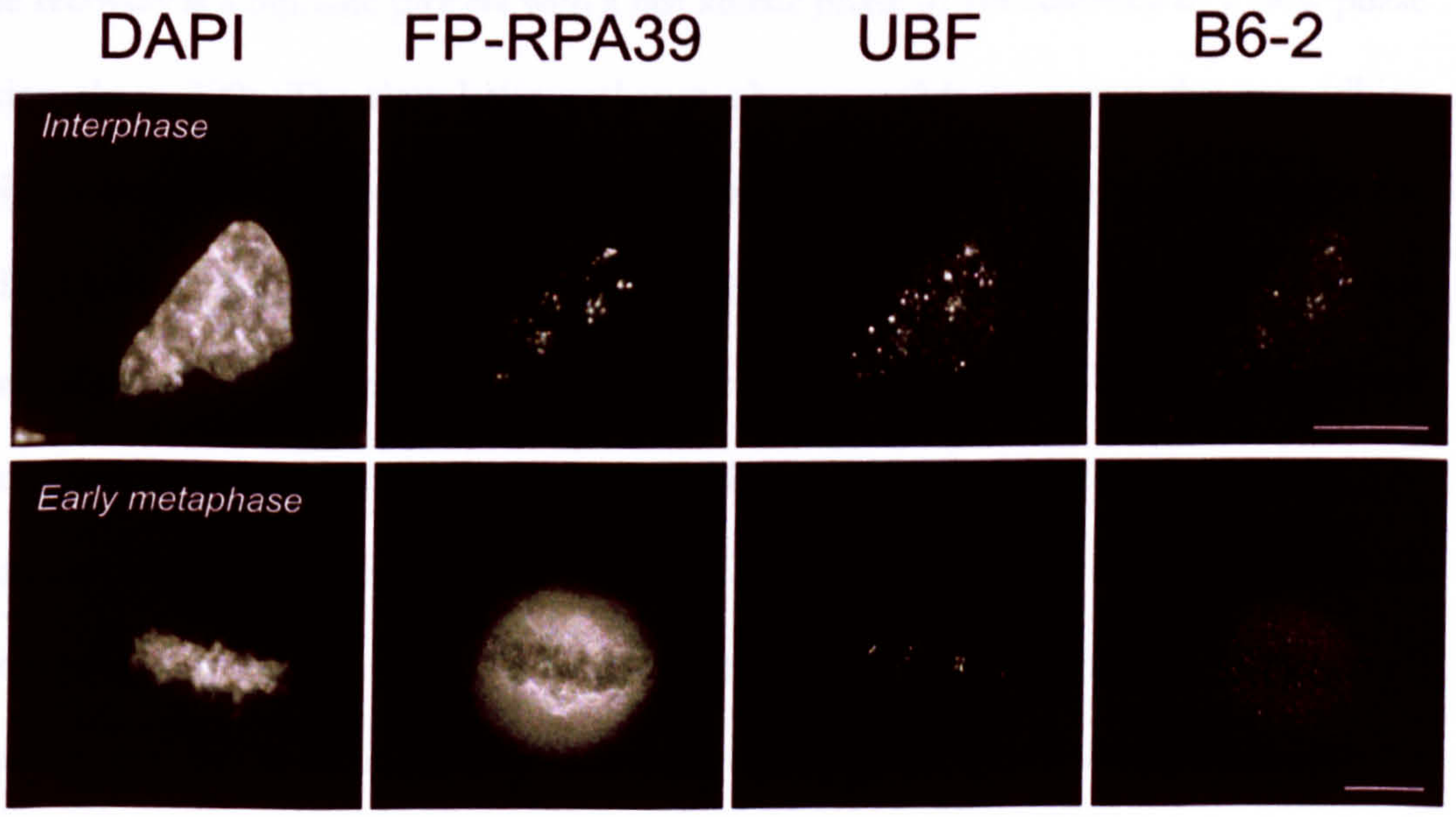
A


00:00
09:03
24:20
34:02
45:54
55:33

B


-10 to -6μm
-6 to 4μm
4-10μm
44:24
45:55
47:30

C

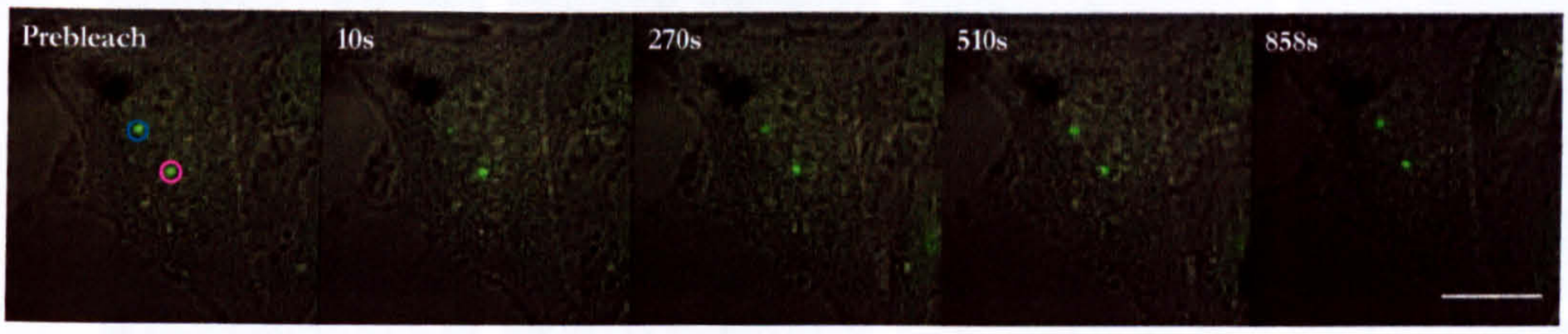

DAPI
FP-RPA39
UBF
B6-2
Interphase
Early metaphase

In summary, when nucleoli break down, UBF remains associated with the chromosomes during all stages of mitosis, while RNA polymerase I subunits leave the chromosomes at a certain point during prophase, after the chromosomes begin to condense. This in turn is followed by the dissolution of the DFC and GC compartments in a short period of time.

IV.4.3 Dynamic properties of a RNA polymerase I subunit

The behaviour of RNA polymerase I subunit RPA39 during mitosis prompted me to investigate the dynamics of recruitment of RNA polymerase I components to endogenous ribosomal genes during interphase using fluorescence recovery after photobleaching (FRAP) analysis (Figure IV-8). Sites of rDNA transcription (FC) in cells expressing the tandem triple EGFPx3-RPA39 were bleached with the use of a short laser pulse that irreversibly quenches the GFP signal. The recovery of the signal in the bleached area (Figure IV-8, upper panel, blue open circle) and the loss of the signal in another FC of a different nucleolus (Figure IV-8, upper panel, red open circle) were recorded by time-lapse confocal microscopy.

The recovery is a biphasic process with a fast kinetic phase of 10s followed by a slow phase taking about 300s. The slow latter exchange phase possibly represents the transcribing RNA polymerases (Figure IV-8, lower panel) whilst the fast exchange may represent the diffusible RNA polymerases at that region. This behaviour was also observed in a recent study of dynamics of other RNA polymerase I subunits (Dundr et al., 2002). Similar results were obtained from another HeLa cell line stably expressing a single ECFP-tagged RPA39 (data not shown) and hence the slow process was not due to the additional copies of fluorescent protein tags in the tandem triple EGFPx3-RPA39 cell line. Moreover, I conclude that the RNA polymerase subunit travels from one FC to another continuously through the nucleoplasm, as indicated by the loss of fluorescence level in the FC of another nucleolus marked by a red open circle, whilst the fluorescence level in the FC



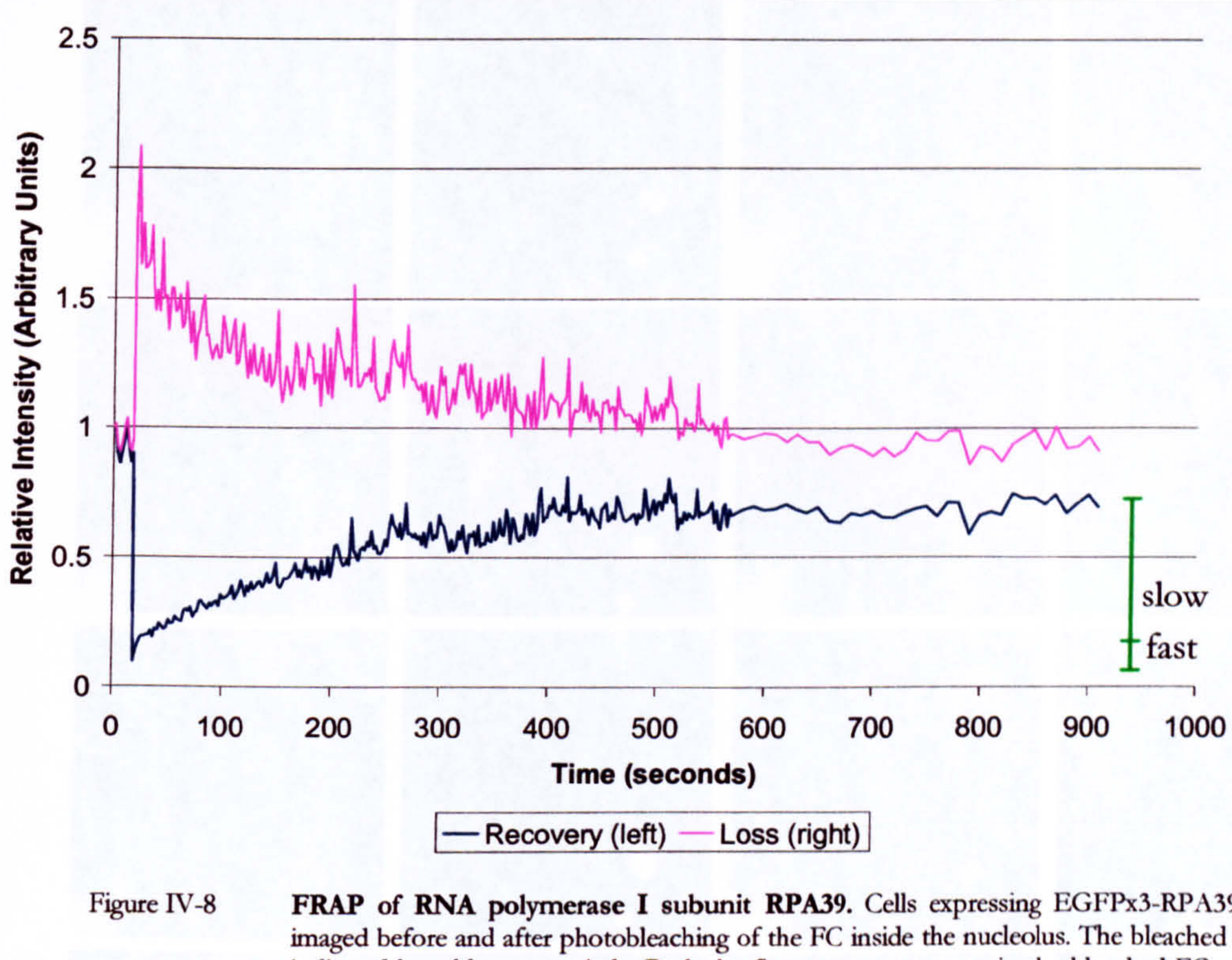


Figure IV-8 **FRAP of RNA polymerase I subunit RPA39.** Cells expressing EGFPx3-RPA39 were imaged before and after photobleaching of the FC inside the nucleolus. The bleached area is indicated by a blue open circle. Both the fluorescence recovery in the bleached FC and the loss of fluorescence in another FC (red open circle) were monitored. The time after photobleaching is indicated on the corresponding upper panel image and the quantitation of recovery kinetics is represented on the lower panel. Relative intensity is defined as the intensity compared with the pre-bleach image after subtracting the background. The initial rise of the fluorescence loss curve for the FC marked by a red open circle is due to the decrease in total fluorescence after photobleaching. Scale bar = 5 μm.

marked in blue recovered after photobleaching. It is apparent that the majority of FP-RPA39 molecules localised in the FC are likely in a transcriptionally active state, and only a small number of them are diffusible (Figure IV-8, lower panel; Dundr et al, 2002). Hence, the disappearance of RNA polymerase I in the FC/NOR (defined by UBF staining in Figure IV-7C) may be explained by the halt of transcription activities during mitosis.

IV.4.4 Establishment of nucleoli

The resumption of transcription activities was inferred by the immediate re-formation of FCs post mitosis (Figure IV-7B & Figure IV-9). Interestingly, although both daughter cells

A

RPA39

FIB

RL27

Merged

00:00

02:09

04:13

06:17

08:22

10:51

33:10

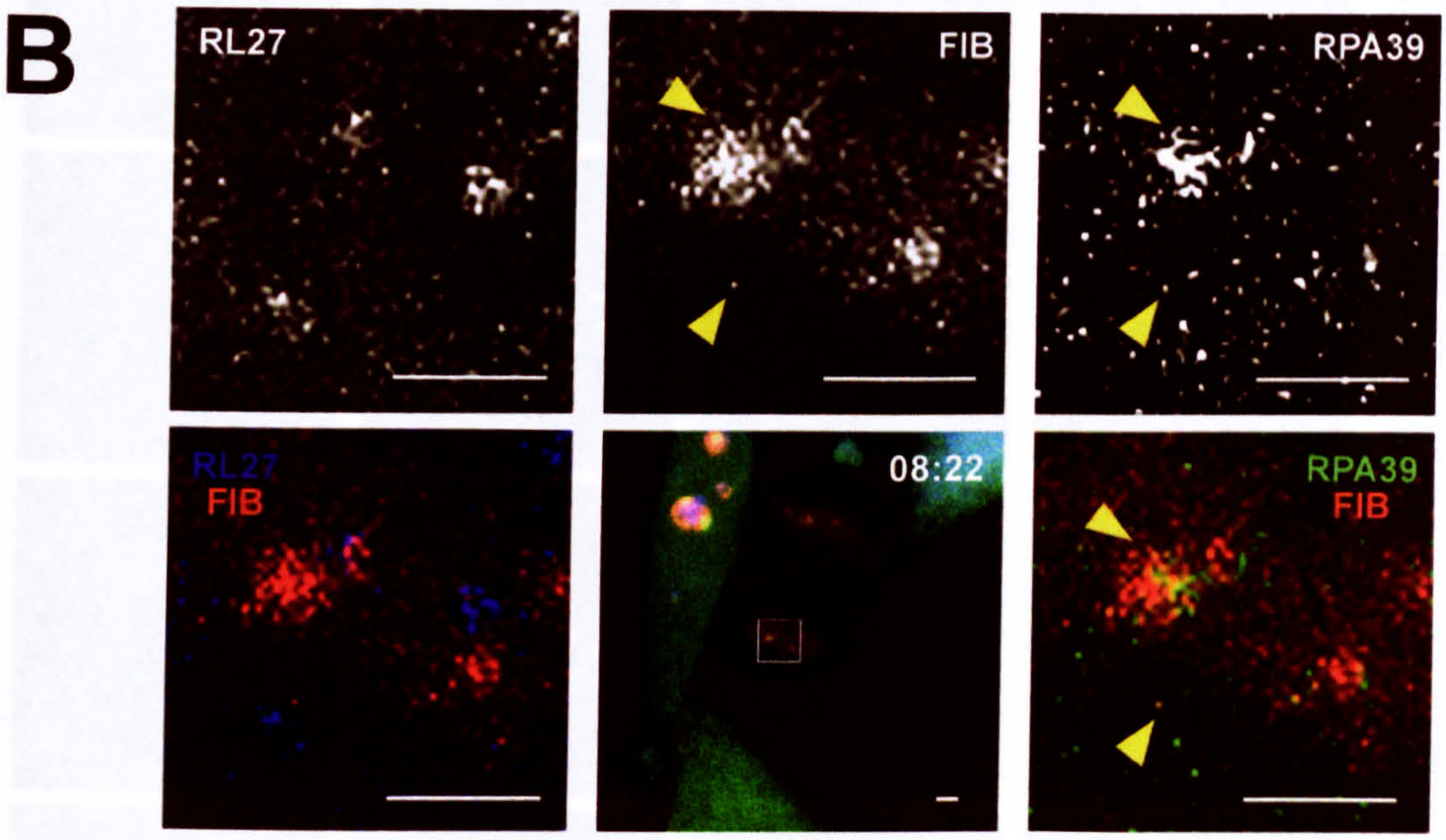


Figure IV-9 **Nucleolar reassembly.** HeLa cells stably expressing both ECFP-FIB and EYFP-RPA39 were transiently transfected with DsRed2-RL27 and imaged after 24 hours. (A) Metaphase cells were located under the microscope using transmission light and images were taken every 2-3 minutes. The series consists of the maximal projections of part of the experiments and the whole series can be found in movie IV-9 in the supplementary CD. Time indicated = min:sec. Scale bar = 5μm. (B) Differential localisations of the DFC and the GC with regard to the FC at an early stage of reassembly from an area denoted by a white square at t=08:22 in panel A. Arrowheads indicate the colocalisation of the RNA polymerase I subunit RPA39 and the RNA processing factor fibrillarin. Scale bar = 1 μm.

received identical genome, not only did the FCs appear at different times in each cell but also the number of FCs were different (Figure IV-9A, first row). It appears therefore that the process of reactivation of rDNA loci and the subsequent formation of FCs is stochastic.

As noted earlier, FCs appeared just after the chromosomes segregated and DFCs appeared roughly 5 minutes later (Figure IV-5, 50:52-60:53 & Figure IV-9A, white arrowheads in first and second row). Fibrillarin relocated mostly around chromosomes as spot like structures once the chromosomes were segregated (Figure IV-9A, 02:09-04:13) and some of them immediately surrounded any available FCs (Figure IV-9B, right half). Previous studies have described the structures represented by fibrillarin at this early stage termed as

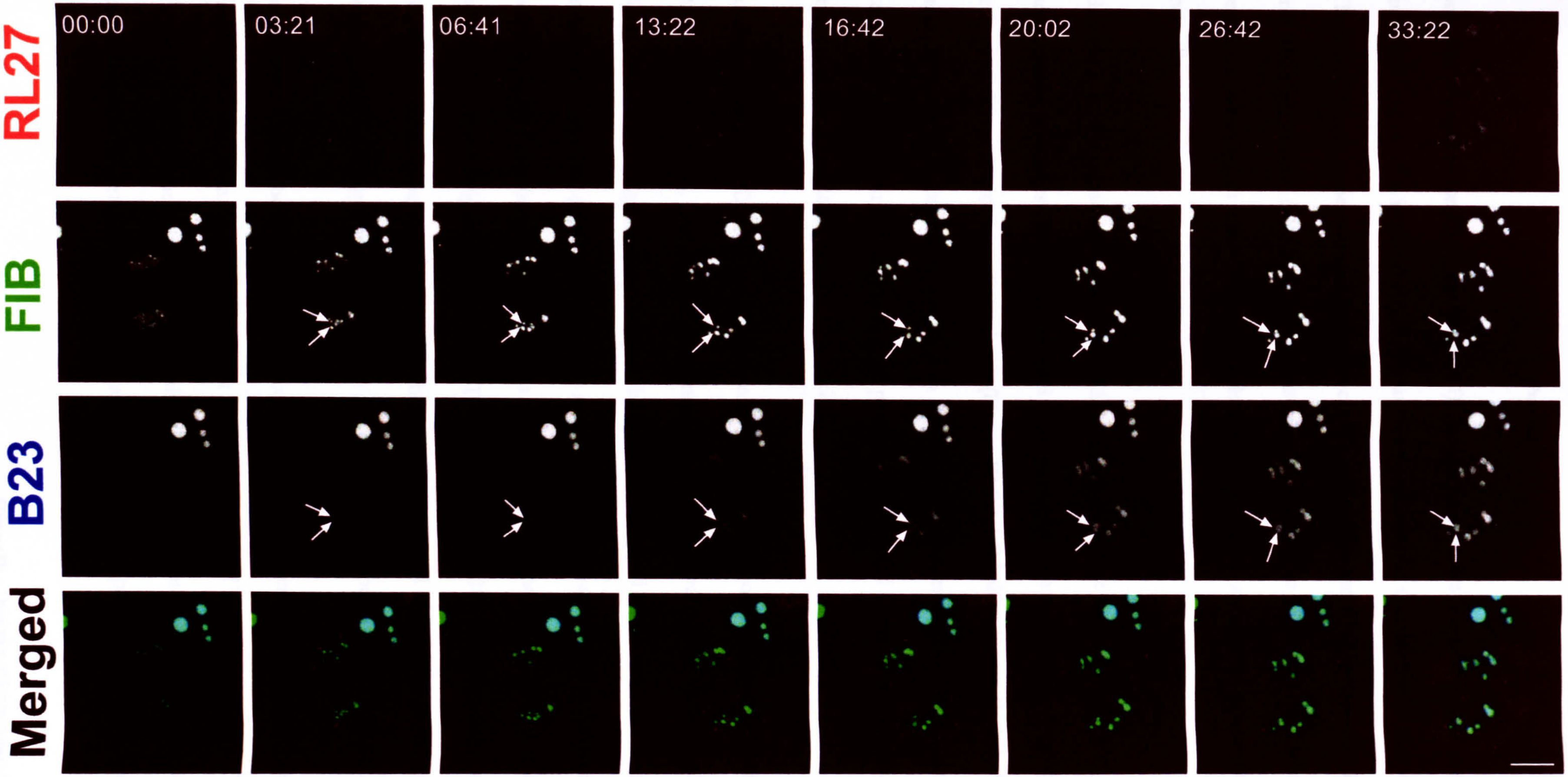


Figure IV-10 **Same genetic material, different structures.** HeLa cells stably expressing both ECFP-B23 and EYFP-FIB were transiently transfected with DsRed2-RL27 and imaged after 24 hours. (A) Metaphase cells were located under the microscope using transmission light and images were taken every 2-3 minutes after anaphase. The series consists of the maximal projection of part of the experiments and the whole series can be found in movie IV-10 in the supplementary CD. White arrows indicate how two DFC merged (see text for details). Time indicated = min:sec. Scale bar = 5μm.

prenucleolar bodies (or PNB; see Chapter I). As previous studies are mostly performed on fixed cell studies, it is impossible to correlate two subnucleolar compartments temporally. Therefore, it may underestimate the connection existing between RNA polymerase I (FC) and RNA processing factors (DFC) at this stage as revealed in Figure IV-9B (right half, yellow arrowheads), in which both subnucleolar compartments are studied concurrently in live cells.

However, at the same time, B23 and RL27 formed discrete bodies away from such prenucleolar structures (defined by FC and DFC markers) even though they were already re-imported into the nucleus after mitosis (Figure IV-9B, left half & Figure IV-10). At this stage, these discrete bodies never fused to the FC/DFC centres (Figure IV-10); yet both non-ribosomal protein B23 and ribosomal protein RL27 located in the same non-nucleolar structures (for example, Figure IV-10 at 33:22).

As the number of FCs increased dramatically after the chromosomes segregated (Figure IV-7 & Figure IV-9), the area of surrounding DFCs increased correspondingly (Figure IV-9). At a certain stage the area occupied by a DFC overlapped with a neighbouring DFC and merged into one (Figure IV-10, white arrows at second row). As the re-formation of nucleoli carried on (at about 25 minutes after anaphase), the GC began to coat the FC/DFC centres (Figure IV-5, Figure IV-9 & Figure IV-10). Consequently, as the FC/DFC centres expanded and merged, the GC area increased and started to merge with neighbouring GCs (e.g. Figure IV-10 white arrows at third row between 16:42-26:42).

Surprisingly, though both daughter cells acquired identical copies of the genome, both the number and size of nucleoli were different from one another (Figure IV-5, Figure IV-9 & Figure IV-10). Whether this is typical for all cells or a specific property of transformed cell lines remains to be determined. Note that although the images presented here are maximal

projections, the difference in the number of nucleoli between the two daughter cells has been checked and verified with the 3D datasets and by DIC images.

IV.4.5 Nucleolar Fusion

Treatment of cells with the RNA polymerase II transcription inhibitor 5,6-dichloro-D-ribofuranosylbenzimidazole (DRB) reversibly causes nucleolar structural disruption and dispersal of DNA components inside the nucleolus, including rDNA and satellite sequences to single ribosomal transcription units (Scheer et al., 1984; Haaf et al., 1991; Junera et al., 1995; Haaf and Ward, 1996). Nucleoli are re-formed after DRB is washed out within 1-2 hours. Therefore, this provides a good experimental window to study in detail how nucleoli re-form *in vivo* (Figure IV-11).

HeLa double stables expressing ECFP-FIB and EYFP-H2B were transiently transfected with diHcRed-H2B and treated with 100μM DRB for 10 hours before imaging. In interphase nuclei of mammalian cells, chromosomes are positioned non-randomly relative to the nuclear periphery according either to gene density or size (see Chapter I). To study how the dispersed nucleolus re-forms, a mark is required for orienting the cells while the nuclei are reorganising within a live cell. The criterion for such a mark is that it should be stable enough not to be affected by cell movements and nuclear rotations. Therefore, FP-tagged histone H2B is a suitable candidate as it is stably associated with chromosomes (about 4 hours), according to the recent FRAP data (Phair and Misteli, 2000; Misteli, 2001). The idea behind this is to label an area by irreversibly photobleaching one type of FP-tagged *H2B* while the photobleached region can be counter-revealed by the same histone, but one tagged with a different FP (Figure IV-11, prebleach; dotted lines surround the photobleached region). This is immediately followed by washing out DRB to observe the re-formation of nucleoli in relation to neighbouring chromatin (Figure IV-11, prebleach).

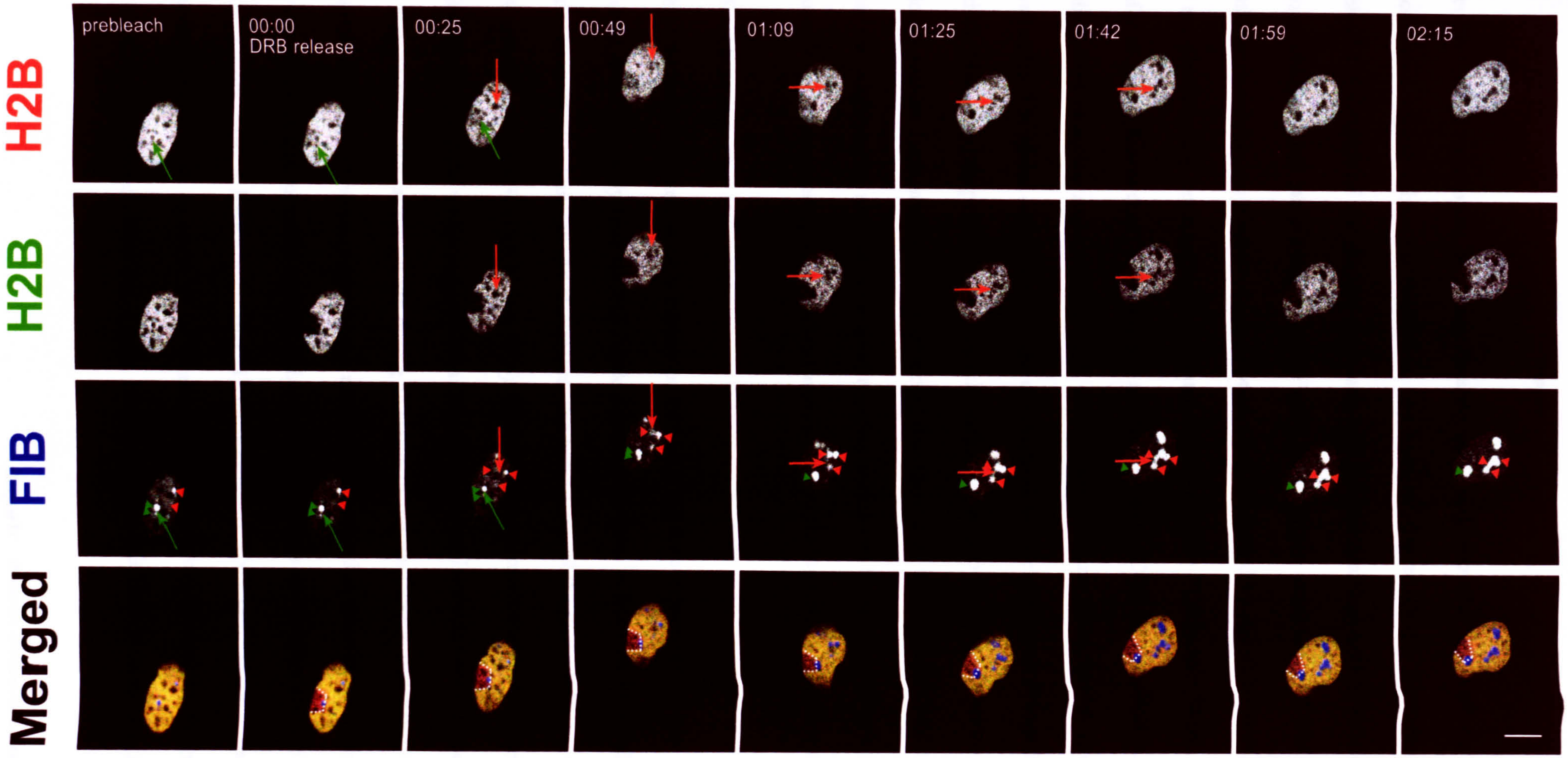


Figure IV-11 **Chromatin and the Nucleolus.** HeLa cells stably expressing both EYFP-FIB and ECFP-H2B were transiently transfected with diHcRed-H2B for 24 hour and treated with 100μM DRB. (The series consists of the maximal projections of the experiment and the whole series can also be found in movie IV-11 in the supplementary CD. Dotted lines surround the photobleached regions, arrowheads indicate how two DFC merged and arrows indicated the interchromosomal space between the fusing nucleoli (see text for details). Time indicated = hr:min. Scale bar = 5μm.

After the DRB treatment, fibrillarin became dispersed into the nucleoplasm with concentrated regions observed throughout the nucleus (Figure IV-11, arrowheads in prebleach). One-fifth of the nuclear chromatin was labelled near two such concentrated regions with a square photobleach mark (Figure IV-11, green arrowheads at 00:00). It should be noted that the ribosomal transcription was not affected by the addition of DRB and the fibrillarin concentrating regions contained one or more FCs (Scheer et al., 1984; Haaf et al., 1991; Junera et al., 1995 and data not shown). Once DRB were washed out, the fibrillarin-concentrating regions fused with one another (Figure IV-11, green and red arrowheads). The global organisation of the chromatin is relatively stable as shown by the relatively unchanged shape of the photobleach mark compared to the reorganisation of the nucleolar factors. Interestingly, I often observed that there were interchromosomal spaces before the neighbouring nucleoli were fused together as deduced from the staining by H2B (Figure IV-11, corresponding green and red arrows indicate the interchromosomal spaces).

IV.5 Discussion

In this chapter, I reported using a battery of 24 HeLa cell lines expressing one or more nucleolar factors to study how nucleoli break down and re-form in live cells by labelling 3 distinct subnucleolar domains and chromatin. The breakdown of the nucleoli began when RNA polymerase I left FCs (c.f. Figure IV-6). This finding apparently contradicts previous studies using fixed cells, which suggested that RNA polymerase I subunits were associated with the chromosomes at the NORs during all stages of mitosis (Scheer and Rose, 1984; Gilbert et al., 1995). However, according to this work, it seems that there exists a window between prophase and metaphase where RNA polymerase I does not localise at the NORs during mitosis. Also, I have shown that (1) the FP-tagged RNA polymerase I subunit behaved properly, both in terms of its localisation at the TEM level and its functional activity at the biochemical level (c.f. Figure IV-2) and (2) confirmed the non-association

with the chromosome during mitosis with antibodies staining another endogenous RNA polymerase I subunit (Figure IV-7c and data not shown). These data correlate well with the FRAP data I have obtained showing that most of the FP-RPA39 molecules localised in FCs are likely actively transcribing (Figure IV-8). The dissociation of RNA polymerase I from chromosomes may provide another level for regulating the transcriptional repression during mitosis. However, as shown by this and other studies (Roussel et al., 1993; Zatsepina et al., 1993; Gebrane-Younes et al., 1997), the transcription factor UBF remained associated with the NORs throughout mitosis. This may be due to its binding to multiple regions in the rDNA locus, including promoters, transcribed sequences and intergenic spacers (O'Sullivan et al., 2002).

The absence of RNA polymerase I in FCs between prophase and metaphase was in turn followed by the immediate breakdown of the DFCs and GCs within a short period of time (c.f. Figure IV-5 & Figure IV-6). The breakdown itself is not the result of chromosome condensation (c.f. Figure IV-5). However, it is observed at a similar period when the nuclear envelope breaks down (Beaudouin et al., 2002; Burke and Ellenberg, 2002) and, in this case, the timing of nuclear envelope breakdown and nucleolar disassembly is strikingly similar. Whether nucleolar disassembly is functionally linked to the nuclear envelope breakdown remains to be determined. Previous studies suggested that the cyclin B-cdk1 kinase, the enzyme partly responsible for nuclear envelope breakdown, is also responsible for phosphorylation of RNA polymerase I specific transcription factors, UBF and SL1, to inhibit mitotic transcription (Klein and Grummt, 1999; Tuan et al., 1999). However, whatever event(s) happened during this period, it must trigger the dissociation of RNA polymerase I complexes from FC and the subsequent DFC/GC breakdown in a short period of time.

Reassembly of nucleoli, in contrast with breakdown, seems to be more hierarchical and the timing of nucleolar re-formation was comparable with how pre-rRNA transcripts progressed through nucleoli during interphase (c.f. Figure IV-4 & Figure IV-9A). It began with the reassociation of RNA polymerase I to the NORs and the rapid increase in number of FCs towards telophase. Whether transcription begins at the same time as the association remains to be established. As soon as the chromosomes segregated, fibrillarin, a component of the DFC, reimported into the re-forming nucleus, appearing as multiple spots. Previous studies have suggested that fibrillarin was localised in a transient structure known as prenucleolar body (PNB) at this stage (Dundr et al., 2000; Savino et al., 2001). The PNB was suggested as a site of temporary storage when nucleolar re-formation was inhibited by the addition of anti-RNA polymerase I antibodies to mitotic cells (Benavente et al., 1987). However, I observed here that many of these spot-like structures were already associated with RNA polymerase I containing sites (c.f. Figure IV-9B). As these RNA polymerase I containing sites are rapidly fusing with one another to form the core of the FC and the fibrillarin containing spots fuse to form the DFC (c.f. Figure IV-9A), it is impossible to predict precisely whether during normal cell cycle progression RNA processing factors like fibrillarin are temporarily stored as defined structures, prior to their association with the FC.

On the other hand, there are discrete non-nucleolar structures containing B23/RL27 that did not associate with the re-forming nucleoli at an early stage of nuclear re-formation (c.f. Figure IV-10). Such structures are similar to those foci that also contained pre-rRNAs synthesised at the G2/M phase of the previous cell cycle (Dousset et al., 2000). It is noted that DRB unravels the compact nucleoli to a necklace-like structure containing components from the FC and the DFC, while GC components disperse into the nucleoplasm, although ribosome biogenesis continue (Scheer et al., 1984). This suggests

that the association of GCs with the FC/DFC regions is not a prerequisite for ribosomal synthesis and it is tempting to speculate that these transient structures may be capable of assembling new ribosomes outside the nucleolus from existing, processed pre-rRNAs.

Recent FRAP analyses have indicated that the nucleolar factors, including RNA polymerase I subunits, transcription factors, RNA processing factors, chaperones and ribosomal proteins, are continually cycling between nucleoli and the nucleoplasm (Dundr et al., 2000; Phair and Misteli, 2000; Chen and Huang, 2001 and also Figure IV-8). The accumulation of these proteins at a specific nuclear structure is most likely due to ongoing molecular interactions. Pre-rRNA transcripts are a likely candidate for mediating such interactions in the re-formation of nucleoli after mitosis. The order of nucleolar re-formation is very similar to the timing of how pre-rRNA transcripts travel from the FC via the DFC to the GC during interphase (5 minutes for the DFC and 30 minutes for the GC; Thiry et al., 2000 and this work). It should be noted that it is the actual rRNA transcripts, rather than the rRNA transcription, that is more important for the re-formation of nucleoli (Dousset et al., 2000; Verheggen et al., 2000). For example, the recruitment of the rRNA processing machineries to the nucleolus is dependent on the recruitment of maternal pre-rRNAs, during *Xenopus laevis* embryogenesis, but independent of either zygotic rRNA transcription, or the presence of RNA polymerase I itself (Verheggen et al., 2000). Similarly, nucleoli can be re-formed from the pre-rRNAs synthesised from the previous cell cycle, even when a low amount of Actinomycin D was added to specifically inhibit RNA polymerase I transcription after mitosis (Dousset et al., 2000). Nevertheless, once transcription resumes after mitosis, the newly made transcripts may make a primary contribution to the physiological re-formation of nucleoli in dividing cells.

But what determines the number and size of nucleoli? From this study, it is surprising to find that the two daughter cells differed in both the number and size of nucleoli after

mitosis (c.f. Figure IV-5, Figure IV-9 & Figure IV-10). If the localisation of RNA polymerase I indicates the site of transcription (Scheer et al., 1984; Haaf et al., 1991), reactivation of transcription seems to take place in different nuclear locations after mitosis in respective daughter cells. This stochastic behaviour and the subsequent re-formation of the DFC/GC around such FCs may partly explain the differences within the pair of daughter cells. On the other hand, an interchromosomal space in between two nucleoli appears to be a prerequisite for the fusion event (Figure IV-11). Recent studies by Chubb et al suggested that the mobility of the chromatin near nucleoli is constrained during interphase. This constraint is released when cells are treated with DRB, but is resumed within 1-4 hours after the washout of DRB (Chubb et al., 2002). The increase in the number of cavities between chromosomes (interchromosomal space; c.f. Figure IV-11, prebleach) may be reflected in the increased chromosomal mobility when DRB is added.

As many nucleolar proteins are continually exchanging between nucleoli and the nucleoplasm, such nucleolar proteins are constantly exploring the available interchromosomal space. Because the resolution of FRAP analyses does not reach the molecular level, the results represent the sum of movement of molecules in all of their possible forms: i.e. bound, free, activated or inactivated. Therefore, the mobility of a substrate-bound complexed form may be overestimated in the current FRAP analyses. The fact that the diffusion rate is at least 10 fold lower in nucleoli than in the nucleoplasm may partly account for the differences in free and substrate-bound forms (Phair and Misteli, 2000) and hence the proteins accumulate to areas where they find their molecular partners to perform specific functions. This may explain what happens when DRB is added to cells, which results in an increase in the available interchromosomal space (Figure IV-11, prebleach). As there is more interchromosmal space for nucleolar proteins to explore, the once compact nucleoli unravel to single transcription units and this results in the typical

appearance of necklace-like structures (Scheer et al., 1984; Haaf et al., 1991). Therefore, the number and size of nucleoli may be determined by the interplay between the interchromosomal space available and molecular diffusion rates of the nucleolar proteins, which decrease as a result of molecular interactions and hence localise nucleolar factors around the substrate, i.e. the pre-rRNA transcript.

Although in this model, 'self-organisation' is likely to have happened during nucleolar re-formation, it does not explain why isolated nucleoli that are transcriptionally active would not fall apart if all components are continually roaming in and out of nucleoli (Cheutin et al., 2002). It is likely that there may be some 'adhesive elements' to maintain the structures. The high abundance of the proteins with RNA binding protein signature motifs RR[GR] (~30%) and RNA recognition motif RRM (~10%) and the protein-protein interaction domains WD (~5%) in the human nucleolar proteome, as deduced from Chapter III, suggests that RNA-protein and protein-protein interactions are crucial and they may play a part in maintaining the structural stability of the nucleolus.

It is interesting to draw a parallel between nucleolar re-formation and the cell sorting mechanisms. When randomly mixed, disaggregated neural cells from frog embryos aggregate into the most compact form, i.e. a sphere, and then sort themselves into a pattern resembling the original tissue, with ectoderm surrounding the endoderm (Townes and Holtfreter, 1955). Many different cells sort similarly and this was explained by their inherent differences in adhesiveness to one another. The random movement creates initial contacts, while only those cells adhering strongly stay together, resulting in the distinctive clustering of cell types (Steinberg, 1964; Armstrong, 1989). Similarly, the nucleolar proteins may roam around the nucleus, by diffusion to find their substrates, while the resulting structures formed are maintained by 'adhesive elements'. The molecular nature of such 'adhesive elements' remains to be characterised.

During the cell cycle, chromosomal mobility was shown to be high during early G1 when chromosomes undergo considerable rearrangements (Heun et al., 2001 and Chubb, J.R. , personal communication). The re-initiation of RNA polymerase I transcription post mitosis generates new centres of substrates at the NOR-containing chromosomes (Figure IV-12). Subsequent localisations of RNA processing factors, chaperones and ribosomal proteins around the different processed forms of rRNA transcripts results in the apparent reassembly of nucleoli. The order of reappearance of different subnucleolar domains is hierarchical because the different processed forms of rRNA transcripts are available at different times as the transcripts are subsequently modified. The differences in nucleolar

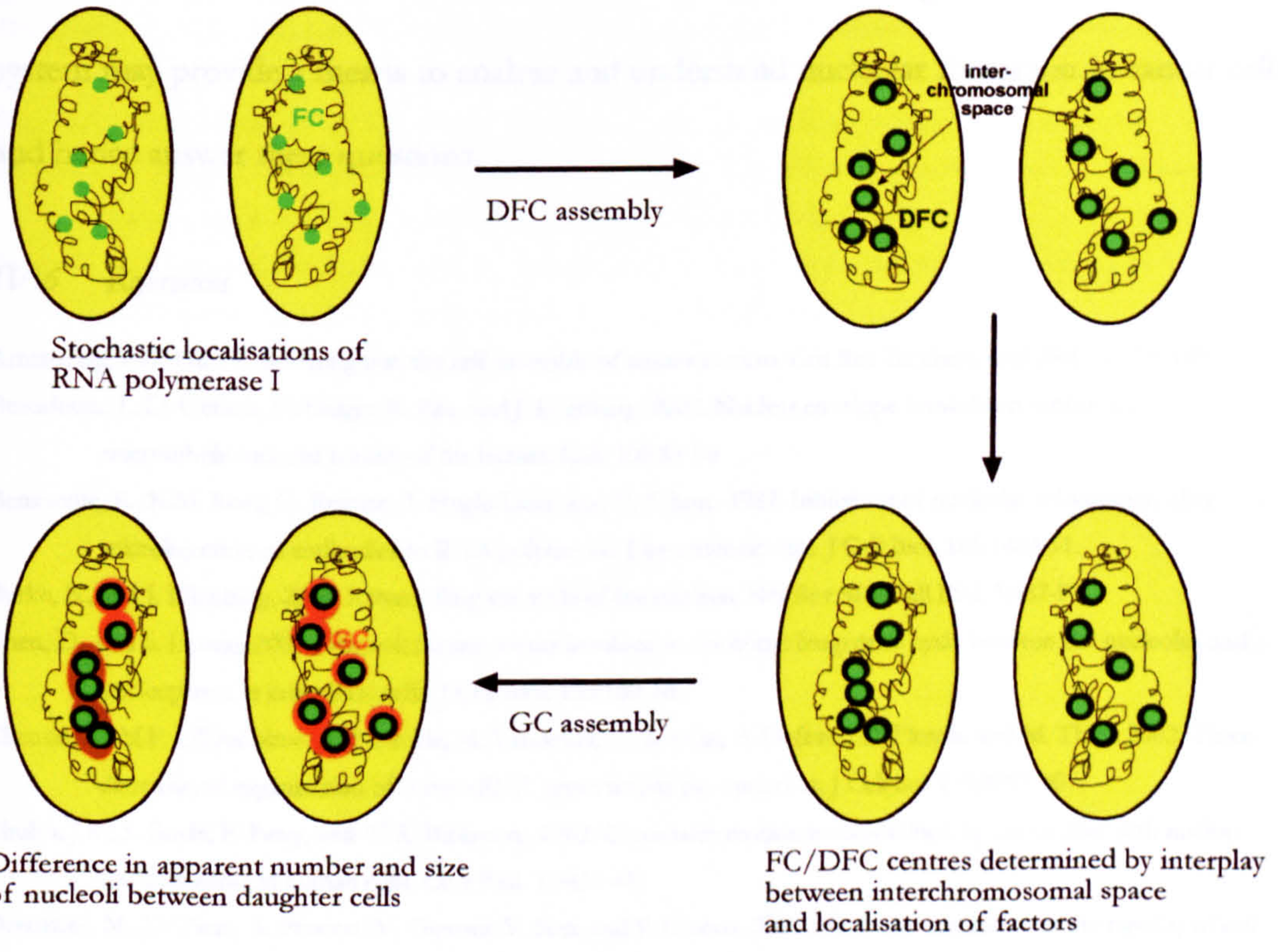


Figure IV-12 **Model of nucleologenesis.** After mitosis, the cells resume RNA transcription as shown by the stochastic localisation of RNA polymerase I within the nucleolus (this study). Around 5 minutes after the pre-rRNA is transcribed, the DFC appears. Though most of the nucleolar factors examined so far constantly exchanged with the nucleoplasm, the factors are accumulated to where the molecular interaction occurs, i.e. around the pre-rRNA transcript. However, the chromosomes present physical barriers for the localisation of nucleolar factors around the substrate and hence this may cause the differences in the number and size of nucleoli even within a pair of daughter cells during early G1.

re-formation in respective daughter cells may be partly explained by slight local differences in the non-random positioning of chromosomes between daughter cells. In other words, the chromatin between two nucleoli may present a temporal, yet physical, barrier for nucleolar fusions. The barriers are temporal because the chromosomes are dynamic and the interchromosomal spaces may change in response to different states of gene expression.

Currently, it remains unclear why the number and size of nucleoli increase dramatically in some malignant cells (Derenzini et al., 2000). Are any of the 400 or more components identified in the nucleolar proteome responsible for the change in the number and size of nucleoli? Or is it the result of the differential state of surrounding chromatin? The current system may provide a means to analyse and understand nucleolar formation in cancer cells and hence answer these questions.

IV.6 References

Armstrong, P.B. 1989. Cell sorting out: the self-assembly of tissues in vitro. Crit Rev Biochem Mol Biol. 24:119-149.

Beaudouin, J., D. Gerlich, N. Daigle, R. Eils, and J. Ellenberg. 2002. Nuclear envelope breakdown proceeds by microtubule-induced tearing of the lamina. Cell. 108:83-96.

Benavente, R., K.M. Rose, G. Reimer, B. Hugle-Dorr, and U. Scheer. 1987. Inhibition of nucleolar reformation after microinjection of antibodies to RNA polymerase I into mitotic cells. J Cell Biol. 105:1483-91.

Burke, B., and J. Ellenberg. 2002. Remodelling the walls of the nucleus. Nat Rev Mol Cell Biol. 3:487-97.

Chen, D., and S. Huang. 2001. Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. J Cell Biol. 153:169-76.

Cheutin, T., M.F. O'Donohue, A. Beorchia, M. Vandelaer, H. Kaplan, B. Defever, D. Ploton, and M. Thiry. 2002. Three-dimensional organization of active rRNA genes within the nucleolus. J Cell Sci. 115:3297-307.

Chubb, J.R., S. Boyle, P. Perry, and W.A. Bickmore. 2002. Chromatin motion is constrained by association with nuclear compartments in human cells. Curr Biol. 12:439-45.

Derenzini, M., D. Trere, A. Pession, M. Govoni, V. Sirri, and P. Chieco. 2000. Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. J Pathol. 191:181-6.

Dousset, T., C. Wang, C. Verheggen, D. Chen, D. Hernandez-Verdun, and S. Huang. 2000. Initiation of nucleolar assembly is independent of RNA polymerase I transcription. Mol Biol Cell. 11:2705-17.

Dundr, M., U. Hoffmann-Rohrer, Q. Hu, I. Grummt, L.I. Rothblum, R.D. Phair, and T. Misteli. 2002. A kinetic framework for a mammalian RNA polymerase in vivo. Science. 298:1623-6.

Dundr, M., T. Misteli, and M.O. Olson. 2000. The dynamics of postmitotic reassembly of the nucleolus. J Cell Biol. 150:433-46.

Gebrane-Younes, J., N. Fomproix, and D. Hernandez-Verdun. 1997. When rDNA transcription is arrested during mitosis, UBF is still associated with non-condensed rDNA. J Cell Sci. 110 (Pt 19):2429-40.

Gilbert, N., L. Lucas, C. Klein, M. Menager, N. Bonnet, and D. Ploton. 1995. Three-dimensional co-location of RNA polymerase I and DNA during interphase and mitosis by confocal microscopy. J Cell Sci. 108 (Pt 1):115-25.

Haaf, T., D.L. Hayman, and M. Schmid. 1991. Quantitative determination of rDNA transcription units in vertebrate cells. Exp Cell Res. 193:78-86.

Haaf, T., and D.C. Ward. 1996. Inhibition of RNA polymerase II transcription causes chromatin decondensation, loss of nucleolar structure, and dispersion of chromosomal domains. Exp Cell Res. 224:163-73.

Heun, P., T. Laroche, K. Shimada, P. Furrer, and S.M. Gasser. 2001. Chromosome dynamics in the yeast interphase nucleus. Science. 294:2181-6.

Huang, S. 2002. Building an efficient factory: where is pre-rRNA synthesized in the nucleolus? J Cell Biol. 157:739-41.

Junera, H.R., C. Masson, G. Geraud, and D. Hernandez-Verdun. 1995. The three-dimensional organization of ribosomal genes and the architecture of the nucleoli vary with G1, S and G2 phases. J Cell Sci. 108 (Pt 11):3427-41.

Klein, J., and I. Grummt. 1999. Cell cycle-dependent regulation of RNA polymerase I transcription: the nucleolar transcription factor UBF is inactive in mitosis and early G1. Proc Natl Acad Sci U S A. 96:6096-101.

Miller, G., K.I. Panov, J.K. Friedrich, L. Trinkle-Mulcahy, A.I. Lamond, and J.C. Zomerdijk. 2001. hRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. Embo J. 20:1373-82.

Misteli, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. Science. 291:843-7.

O'Sullivan, A.C., G.J. Sullivan, and B. McStay. 2002. UBF binding in vivo is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. Mol Cell Biol. 22:657-68.

Phair, R.D., and T. Misteli. 2000. High mobility of proteins in the mammalian cell nucleus. Nature. 404:604-9.

Roussel, P., C. Andre, C. Masson, G. Geraud, and D. Hernandez-Verdun. 1993. Localization of the RNA polymerase I transcription factor hUBF during the cell cycle. J Cell Sci. 104 (Pt 2):327-37.

Savino, T.M., J. Gebrane-Younes, J. De Mey, J.B. Sibarita, and D. Hernandez-Verdun. 2001. Nucleolar assembly of the rRNA processing machinery in living cells. J Cell Biol. 153:1097-110.

Scheer, U., B. Hugle, R. Hazan, and K.M. Rose. 1984. Drug-induced dispersal of transcribed rRNA genes and transcriptional products: immunolocalization and silver staining of different nucleolar components in rat cells treated with 5,6-dichloro-beta-D-ribofuranosylbenzimidazole. J Cell Biol. 99:672-9.

Scheer, U., and K.M. Rose. 1984. Localization of RNA polymerase I in interphase cells and mitotic chromosomes by light and electron microscopic immunocytochemistry. Proc Natl Acad Sci U S A. 81:1431-5.

Shaw, P.J., and E.G. Jordan. 1995. The nucleolus. Annu Rev Cell Dev Biol. 11:93-121.

Steinberg, M.S. 1964. The problem of adhesive selectivity in cellular interactions. In Cellular Membranes in Development. M. Locke, editor. Academic Press, New York. 321-366.

Thiry, M., T. Cheutin, M.F. O'Donohue, H. Kaplan, and D. Ploton. 2000. Dynamics and three-dimensional localization of ribosomal RNA within the nucleolus. Rna. 6:1750-61.

Townes, P.L., and J. Holtfreter. 1955. Directed movements and selective adhesion of embryonic amphibian cells. J Exp Zool. 128:53-120.

Tuan, J.C., W. Zhai, and L. Comai. 1999. Recruitment of TATA-binding protein-TAFI complex SL1 to the human ribosomal DNA promoter is mediated by the carboxy-terminal activation domain of upstream binding factor (UBF) and is regulated by UBF phosphorylation. Mol Cell Biol. 19:2872-9.

Verheggen, C., G. Almouzni, and D. Hernandez-Verdun. 2000. The ribosomal RNA processing machinery is recruited to the nucleolar domain before RNA polymerase I during Xenopus laevis development. J Cell Biol. 149:293-306.

Zatsepina, O.V., R. Voit, I. Grummt, H. Spring, M.V. Semenov, and M.F. Trendelenburg. 1993. The RNA polymerase I-specific transcription initiation factor UBF is associated with transcriptionally active and inactive ribosomal genes. Chromosoma. 102:599-611.

CHAPTER V

A NOVEL NUCLEOLAR TARGETING PATHWAY

"Next Stop, Nucleolus!"

~ The title in Journal of cell biology cover, 13th May 2002~

V. A Novel Nucleolar Targeting Pathway

V.1 Introduction

The cell nucleus is the site where chromosomes are located and where DNA replication and gene expression are coordinated and regulated. Many nuclear factors are organized into specific structures termed "nuclear bodies/domains" (See Chapter I). Most subnuclear structures are dynamic, disassembling on entry to M phase and reassembling after mitosis as illustrated in the previous chapter. Unlike cytoplasmic organelles, subnuclear structures are not enclosed by membranes. Factors can move in and out of subnuclear structures (Misteli, 2001); and the bodies themselves can also move within the nucleoplasm (Boudonck et al., 1999; Platani et al., 2000; Snaar et al., 2000; Muratani et al., 2001). So, how do proteins know where to localise?

The nucleus is separated by the nuclear envelope from the cytoplasm, where translation occurs (Burke and Ellenberg, 2002; Hetzer et al., 2002). To enter nuclei, proteins containing nuclear localisation signals (NLS, usually one or more stretch of basic amino acids) are actively transported inward through the nuclear pore complexes by nuclear import receptors (Weis, 2002). The GTPase Ran creates a GTP/GDP gradient across the nuclear membrane during interphase and provides the directionality for nuclear transport. However, once proteins get into nuclei, how they localise to particular subnuclear structure remains unclear. Previous studies have shown that there are several nucleolar targeting signals identified (Hatanaka, 1990). The signals were shown to direct exogenous protein to nucleoli and are essential for function. For example, the proper localisation signal of adenovirus protein V, KKEEQDYKPRKLKRVKKKKK, is essential for viral replication (Okuwaki et al., 2001). However, the mechanism behind how these signals function remained unclear.

The advent of green fluorescent protein (GFP) technology allows the visualisation of tagged proteins and their movement in living cells (Swedlow and Lamond, 2001). The use of photobleaching techniques with the GFP fusion proteins also allows analysis of protein dynamics (Phair and Misteli, 2001; Reits and Neefjes, 2001). In this chapter, I describe a novel nucleolar targeting pathway shown by the nucleolar factor NHPX, in which NHPX accumulates in nucleoli and show that it involves an unexpected, transient interaction with splicing speckles. The movement of NHPX from speckles to nucleoli is dependent on pre-mRNA transcription. These data suggest that NHPX may be involved in, and possibly link, several parallel RNA metabolic pathways that occur in distinct nuclear domains.

V.2 In vivo localisation of NHPX

V.2.1 NHPX is a highly conserved protein required in more than one RNA machinery

The NHPX protein is highly conserved from human to plant (Figure V-1). It was originally identified as a putative human homologue of yeast NHP2p (Saito et al., 1996) and later independently as a 15.5kDa RNA-binding protein (Nottrott et al., 1999). Subsequent analyses indicated that NHPX is the human orthologue of the *Saccharomyces cerevisiae* protein

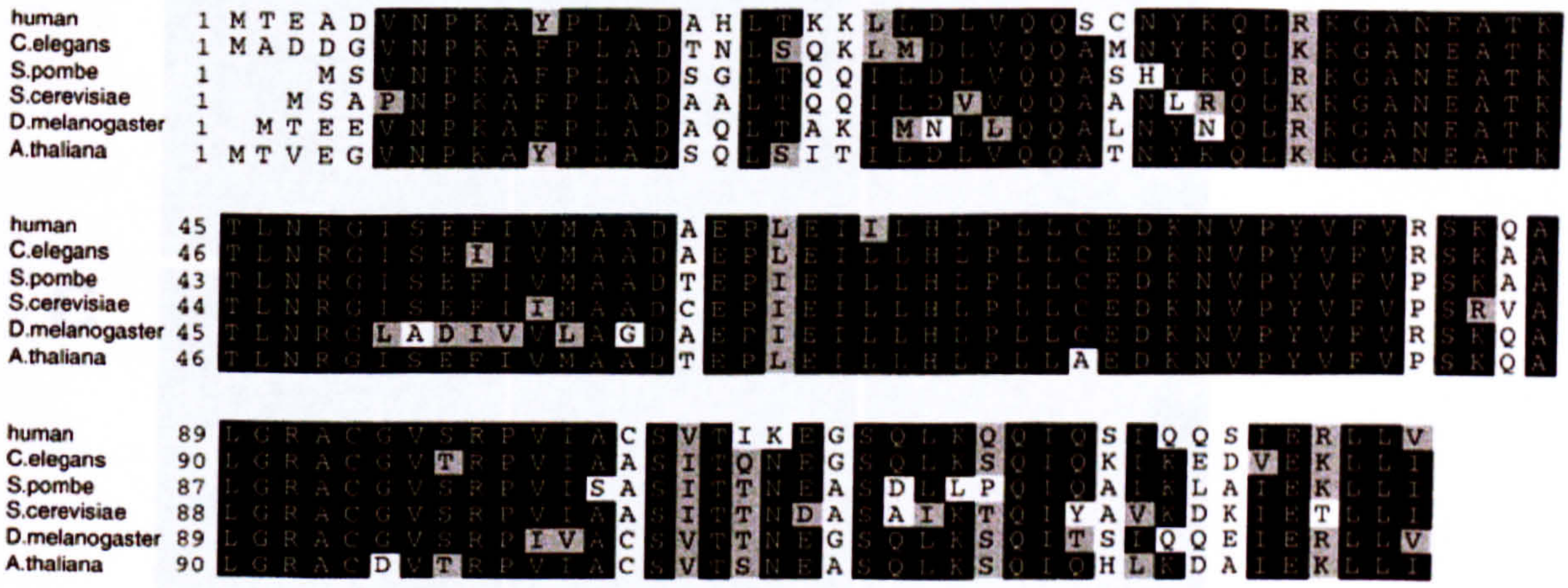


Figure V-1 **NHPX is highly conserved from human to plant**. Protein sequences from *H. sapiens, C.elegans, S. pombe, S. cerevisiae, D. melanogaster* and *A. thaliana* were aligned using Clustal X software. The black boxes indicate identical residues while grey boxes indicate conserved substitutions.

Snu13p (Chang et al., 1999; Nottrott et al., 1999). It shares a common RNP structure that binds to both box C/D snoRNAs and spliceosomal U4 snRNA (Nottrott et al., 1999; Vidovic et al., 2000; Watkins et al., 2000). NHPX was selected for characterisation as part of the analysis of the human nucleolar proteome (Andersen et al., 2002).

V.2.2 NHPX localises in Nucleoli and Cajal bodies

To address its *in vivo* localisation, the NHPX cDNA was isolated from a HeLa cDNA library and tagged at its amino-terminus with the enhanced yellow fluorescent protein (EYFP) gene to generate plasmid pAL107$^{EYFP\text{-}NHPX}$ (see Chapter II). The localisation of the EYFP-NHPX fusion protein was analyzed by fluorescence microscopy following transient transfection of pAL107EYFP-NHPX in HeLa cells (Figure V-2), and compared with the

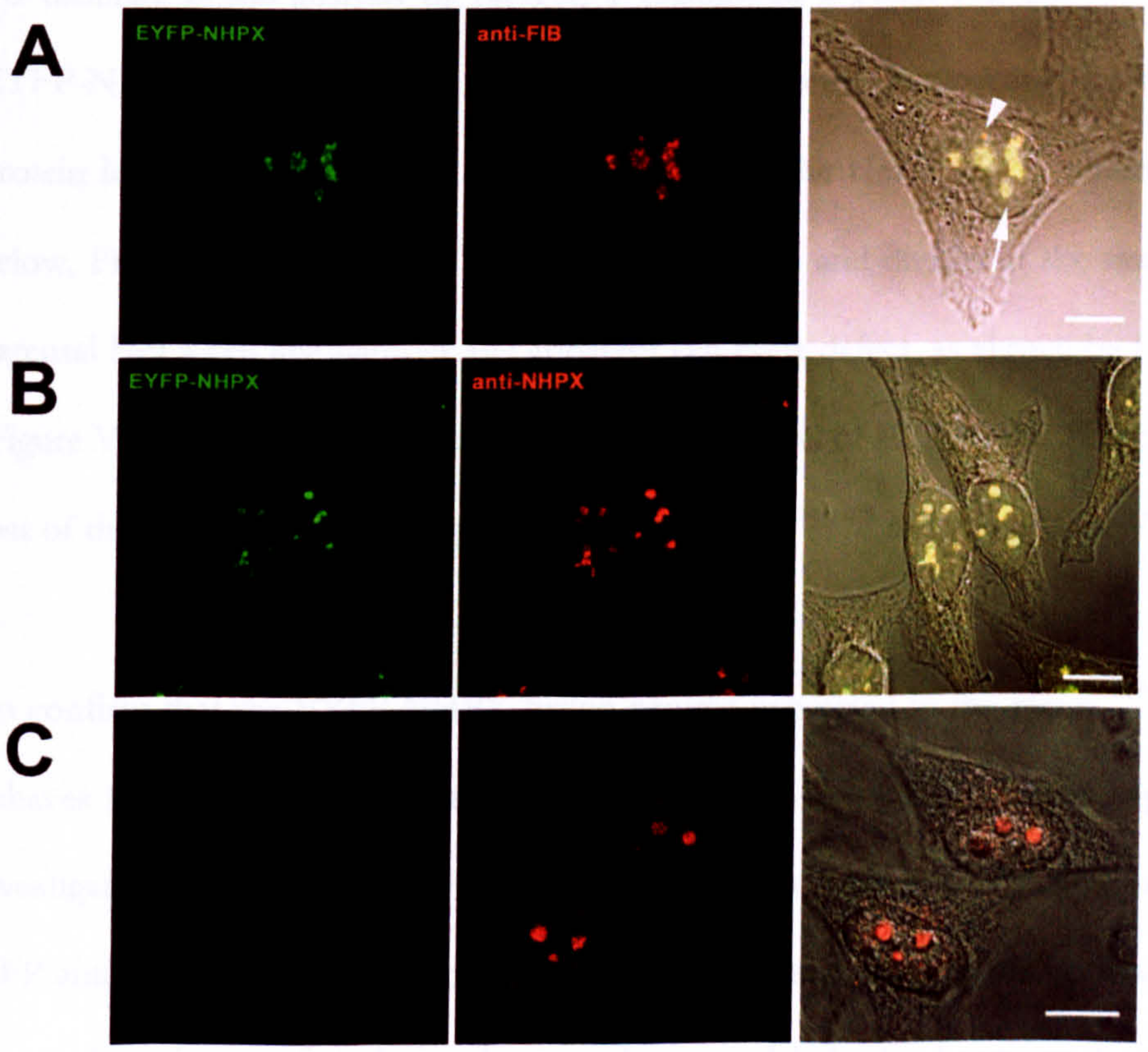


Figure V-2 **NHPX localised in Nucleoli and Cajal bodies in living cells.** HeLa cells transiently transfected with pAL107$^{EYFP\text{-}NHPX}$ for 16 hours were fixed and counterstained with (A) anti-fibrillarin antibody 72B9 and arrowhead indicates CB and arrow indicates nucleolus. Immunofluorescence labeling of HeLa cells using anti-NHPX antibodies either (B) with or (C) without transient transfection with pAL107$^{EYFP\text{-}NHPX}$. Scale Bar = 5μm

nucleolar protein fibrillarin (FIB; Figure V-2A). Transiently expressed EYFP-NHPX was colocalised with fibrillarin in the dense fibrillar component of nucleoli and also, unexpectedly, in Cajal bodies (CBs; Figure V-2A; arrow indicates nucleolus and arrowhead indicates CB). The CB localisation was confirmed by counterstaining with anti-coilin antibodies (data not shown, see Figure V-4A). The localisation of NHPX to CBs, as well as nucleoli, was also confirmed by immunofluorescence using anti-NHPX antibodies (Figure V-2B & C and data not shown). The tagged NHPX showed an identical localisation to its endogenous counterpart, suggesting that NHPX is a component of both nuclear structures (Figure V-2B & C).

V.2.3 Characterisation of a HeLa cell line stably expressing EYFP-NHPX

To facilitate *in vivo* analysis of NHPX, I established a HeLa cell line stably expressing EYFP-NHPX (see Chapter II). As shown by fluorescence microscopy, the EYFP-NHPX protein localised specifically in nucleoli and CBs in the HeLa$^{EYFP\text{-}NHPX}$ stable cell line (see below, Figure V-4). The HeLa$^{EYFP\text{-}NHPX}$ cell line grows and divides at the same rate as the parental HeLa cell line without any apparent cell cycle defect, as shown by FACS analysis (Figure V-3A and data not shown). The expression level of EYFP-NHPX is comparable to that of the endogenous NHPX protein in the HeLa$^{EYFP\text{-}NHPX}$ cell line (Figure V-3B).

To confirm that the EYFP-NHPX fusion protein expressed in the HeLa$^{EYFP\text{-}NHPX}$ cell line behaves biochemically as the endogenous protein, its *in vivo* RNA-binding specificity was investigated. An extract from HeLa$^{EYFP\text{-}NHPX}$ cells was immunoprecipitated using an anti-GFP antibody. RNAs in the immunoprecipitate were separated by urea-polyacrylamide gel electrophoresis, transferred to nylon membrane and hybridised with probes for U1, U2, U4, U5 and U6 snRNAs and U3 snoRNA (Figure V-3C). This showed that U3 snoRNA and U4 snRNA, but not U1, U2, U5 or U6 snRNAs, were preferentially isolated with the

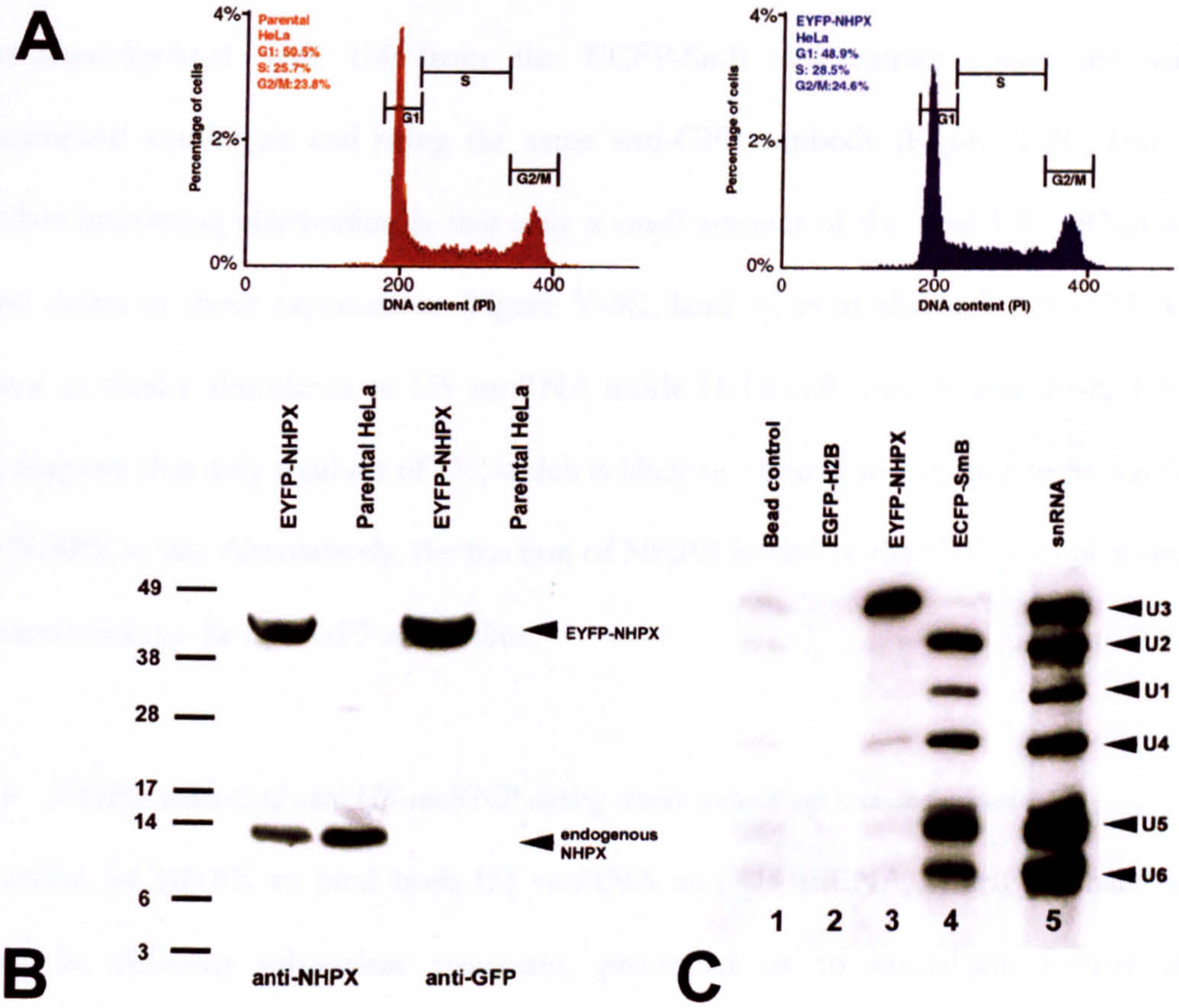


Figure V-3 **Characterisation of HeLa$^{EYFP-NHPX}$ stable cell line.** (A) DNA content of HeLa$^{EYFP-NHPX}$ cells was analyzed by FACS analysis and (B) its expression level by immunoblot using antiserum R86 and anti-GFP. (C) The *in vivo* RNA binding activity of EYFP-NHPX was assayed by immunoprecipitation and the binding of snRNAs U1, U2, U4, U5 and U6 as well as snoRNA U3 were tested by northern hybridisation.

anti-GFP antibody (lane 3). Control experiments, i.e. bead control and an HeLa$^{EGFP-H2B}$ cell extract, showed that although the anti-GFP antibody still pulled down the FP-fusion protein, it did not pull down any of the RNAs tested from these extracts (lane 1 and 2), while the same anti-GFP antibody pulled down each of the U1, U2, U4, U5 and U6 snRNAs, but not U3 snoRNA, from a HeLa$^{ECFP-SmB}$ cell extract (lane 4). These data show that the EYFP-NHPX fusion protein *in vivo* is specifically complexed with the same RNA targets that NHPX was shown to bind directly *in vitro* (Nottrott et al., 1999; Watkins et al., 2000). Given that U4 and U6 snRNAs usually exist as a duplex inside the nucleus, the

immunoprecipitation of U4 snRNA alone is surprising. The disruption of the U4/U6 interaction in this assay is unlikely, because a non-Sm RNA, i.e. U6 snRNA, can be co-immunoprecipitated with U4 from the ECFP-SmB cell extract under the same experimental conditions and using the same anti-GFP antibody (Figure V-3C, lane 4). Another interesting observation is that only a small amount of the total U4 snRNA was pulled down in these experiments (Figure V-3C, land 3), even although U4 snRNA is present in similar abundance to U3 snoRNA inside HeLa cells (Reddy and Bush, 1988). This suggests that only a subset of U4, which is likely not bound to U6, may be interacting with NHPX *in vivo*. Alternatively, the fraction of NHPX bound to the U4/U6 duplex could be inaccessible to the anti-GFP antibodies.

V.2.4 NHPX colocalised with U3-snoRNP during steady state interphase and mitosis

The ability of NHPX to bind both U3 snoRNA and U4 snRNA, which normally are located in different subnuclear structures, prompted us to investigate further the localisation of NHPX under different conditions (Figure V-4). The binding of NHPX to the spliceosomal U4 snRNA suggests that it should colocalise, at least in part, with other splicing factors. However, EYFP-NHPX in the HeLa$^{EYFP\text{-}NHPX}$ cell line does not show a speckled nuclear pattern similar to other human splicing factors such as SC35 and U1A (Figure V-4A; arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles). The anti-NHPX antiserum also does not label speckles (Figure V-2B, Chang et al., 1999). Instead, EYFP-NHPX is accumulated in nucleoli and CBs and colocalises with the snoRNP protein fibrillarin (Figure V-4A and C; see also Figure V-2A). Also, nucleolar segregation, caused by treating the cells with the transcription inhibitor actinomycin D, results in the relocation of NHPX to the nucleolar periphery along with fibrillarin, but does not cause it to colocalise with either SC35, or other splicing factors (see below Figure V-14B and data not shown).

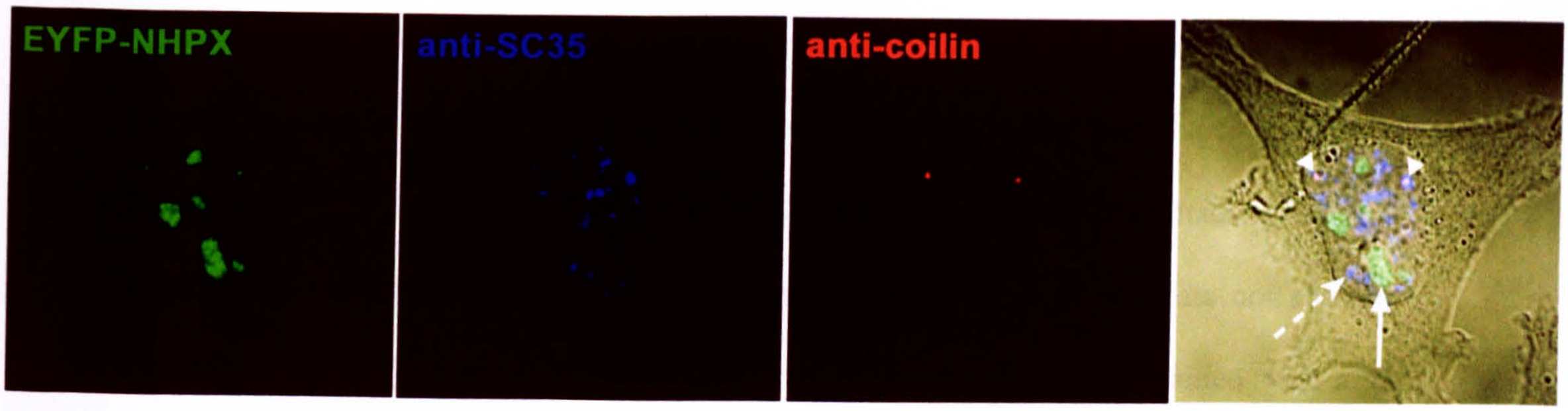


Figure V-4 **The patterns of EYFP-NHPX in HeLa$^{EYFP\text{-}NHPX}$ were analyzed (A) during interphase and (B, C) in different stages of mitosis**. Anti-coilin 204/10 was used to denote CBs, anti-SC35 to denote speckles and DAPI to show the condensed chromosome in mitotic stages. Arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles. (C) Telophase cells were stained with anti-Fibrillarin 72B9 antibodies. (D) The pattern of EYFP-NHPX in HeLa$^{EYFP\text{-}NHPX}$ during mitosis was followed after metaphase in a live cell imaged every 3 minutes. Scale bar=5μm.

Analysis of the HeLa$^{EYFP\text{-}NHPX}$ cells at different stages of mitosis also showed no evidence for the association of NHPX with splicing factors (Figure V-4B). Time-lapse analysis of live HeLa$^{EYFP\text{-}NHPX}$ cells showed that NHPX associated with snoRNP-containing nucleoli immediately after the nuclear envelope reforms (Figure V-4C & D and data not shown). Combined with other immunofluorescence data of fixed cells counterstained with anti-fibrillarin (Figure V-4C & D), I conclude that EYFP-NHPX does not accumulate with snRNPs in speckles, but instead colocalises with the snoRNP protein fibrillarin at all cell cycle stages and metabolic conditions tested. The only colocalisation of NHPX with splicing snRNPs detected *in vivo* was specifically in CBs, structures known to accumulate newly imported snRNPs and snoRNPs when they first enter the nucleus.

V.3 A novel pathway involving transient interaction with splicing speckles

V.3.1 Evidence from microinjection of a plasmid expressing EYFP-NHPX

I next investigated the localisation of newly synthesised NHPX because recent data have shown a temporal pathway for snRNP and snoRNP localisation in nuclei (Carvalho et al., 1999; Narayanan et al., 1999a; Narayanan et al., 1999b; Sleeman and Lamond, 1999; Sleeman et al., 2001). To our surprise, microinjection of the expression vector pAL107$^{EYFP\text{-}NHPX}$ into the cultured cell lines, followed by examination in the fluorescence microscope, revealed that 1 hour after microinjection EYFP-NHPX is accumulated in splicing speckles and CBs (Figure V-5 and see below Figure V-11). 2-6 hours post-microinjection, EYFP-NHPX is also detected in nucleoli, whereas the level of EYFP-NHPX in speckles shows a concomitant decrease. At later time points, the signal in speckles can no longer be detected and EYFP-NHPX accumulates specifically in nucleoli and CBs, resembling the pattern observed in the HeLa$^{EYFP\text{-}NHPX}$ cell line during interphase.

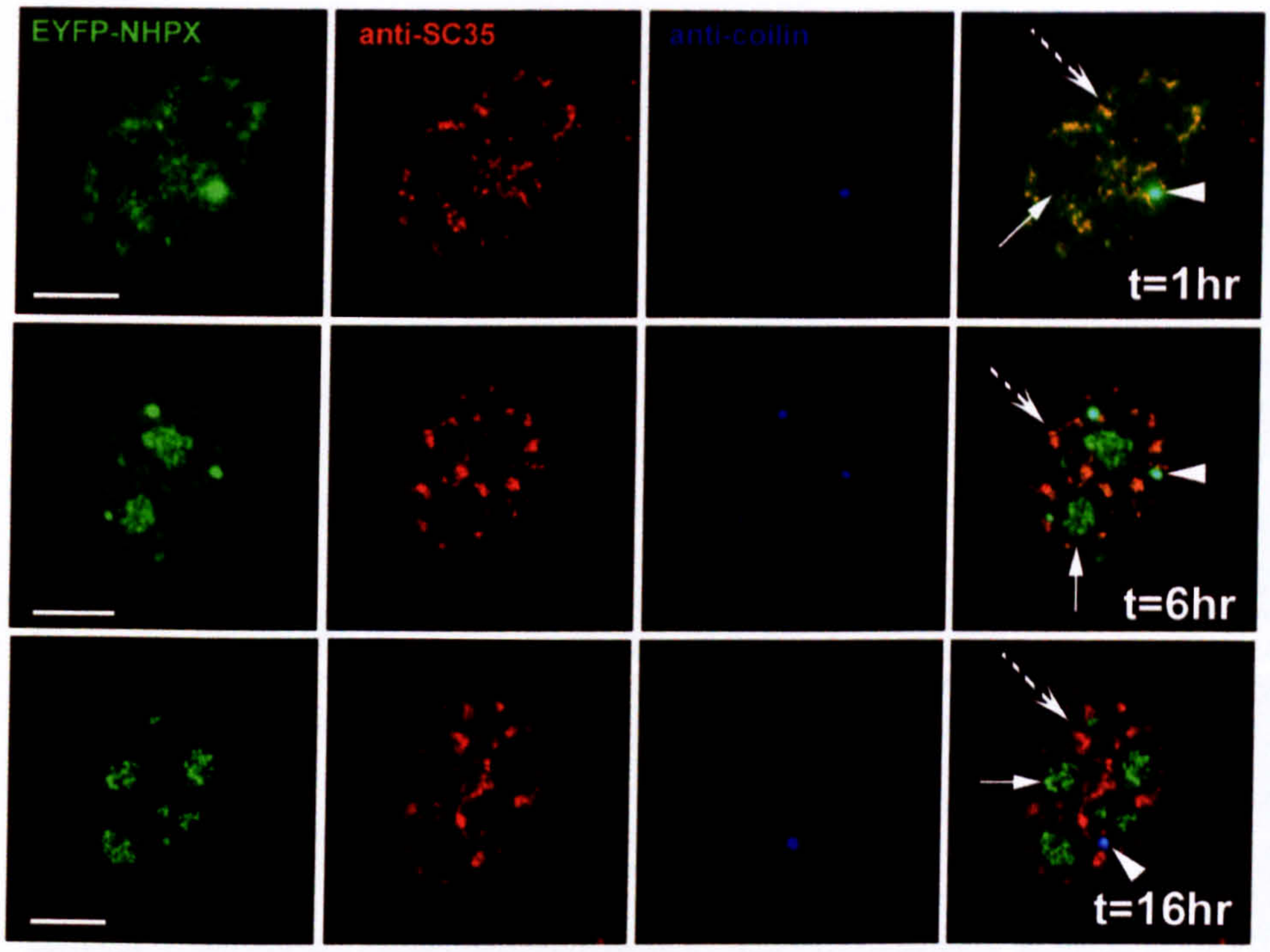


Figure V-5 **Microinjection of pAL107$^{EYFP\text{-}NHPX}$ into transformed cell line MCF7.** The microinjected cells were fixed at different time points (1hr, 6hr and 16hr) and counterstained with anti-SC35 to denote speckles and anti-coilin 204/10 to denote CBs. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles. Scale bar=5μm.

This novel nuclear pathway for NHPX was observed not only in parental HeLa cells (see below) but also when newly synthesised EYFP-NHPX was expressed following either microinjection, or transient transfection, in other transformed cell lines, including MCF7 (Figure V-5; arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles) and HEK293 and in primary cell lines, e.g. human foreskin fibroblasts and primary human fibroblasts expressing telomerase htert1787 (Figure V-6 and other data not shown). Some cell lines, e.g. htert1787, do not contain prominent CBs and therefore provide an opportunity to test whether the presence of CBs is required for the transport and/or localisation of NHPX in nucleoli. Microinjection of pAL107$^{EYFP\text{-}NHPX}$ into htert1787 cells did not induce the formation of CBs and EYFP-NHPX still relocated in the same temporal sequence from speckles to nucleoli (Figure V-6; arrow indicates nucleolus and broken arrow indicates speckles). Therefore, the presence of CBs is apparently not required for the directional movement of NHPX from speckles to nucleoli.

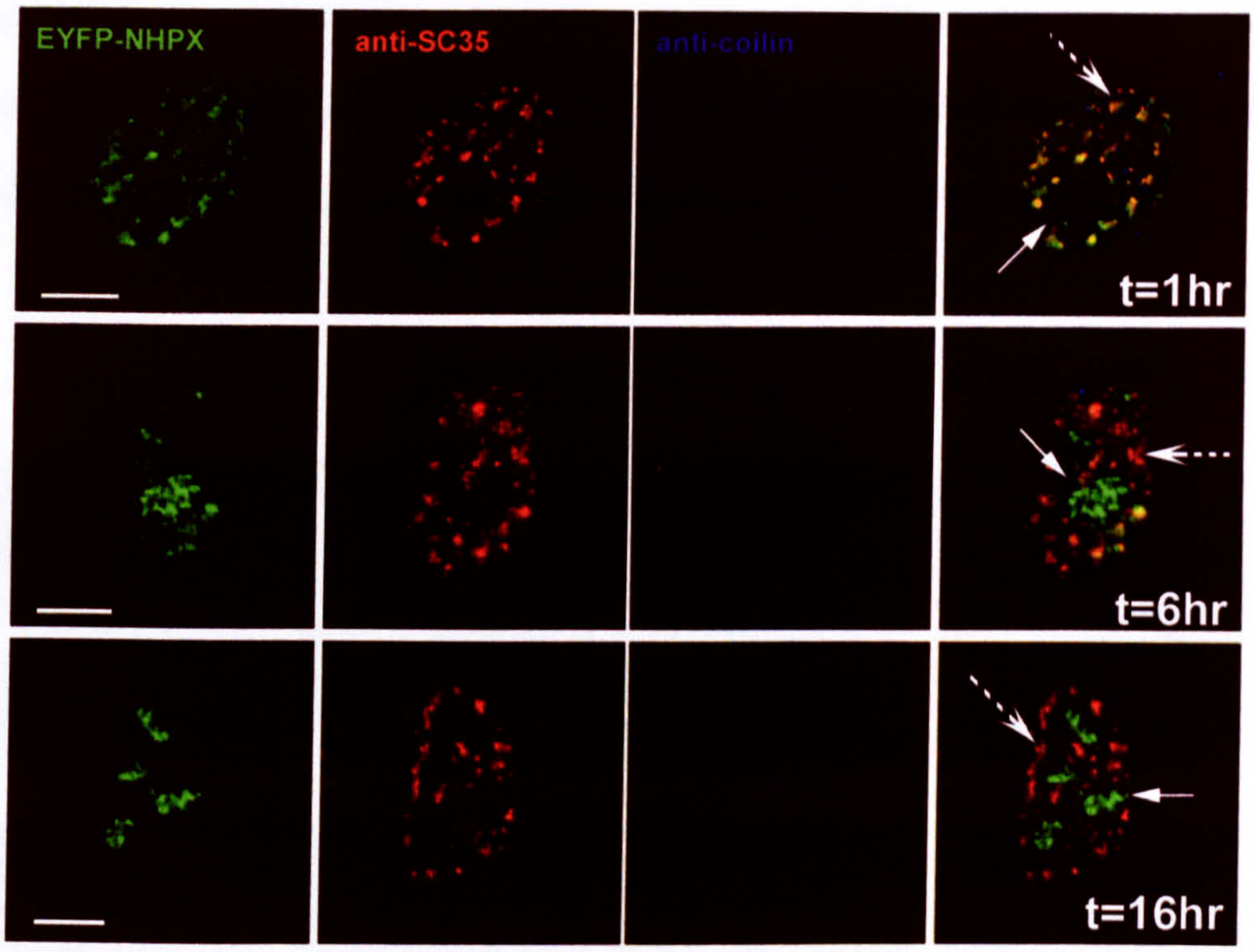


Figure V-6 **Microinjection of pAL107$^{EYFP\text{-}NHPX}$ into primary cell line htert1787**. The microinjected cells were fixed at different time points (1hr, 6hr and 16hr) and counterstained with anti-SC35 to denote speckles and anti-coilin 204/10 to denote CBs. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles. Scale bar=5μm.

The expression of exogenous NHPX raises the possibility that the target RNAs it binds to may be upregulated and thereby results in the observed temporal sequence of nuclear localisation. Therefore, I investigated the *in vivo* localisations of U3 and U4 RNAs in the HeLa cells that were microinjected with pAL107$^{EYFP\text{-}NHPX}$ using complementary 2'-*O*-methyl antisense RNAs (Figure V-7; reviewed in Lamond, 1993). The U3 in microinjected cells remained localised in the dense fibrillar component of nucleoli and CBs at both early and late time points, similar to control, non-microinjected cells (Figure V-7A and B, respectively, arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles). The U4 snRNA was detected in speckles and CBs in both the microinjected cells at different time points and in control, non-microinjected cells (Figure V-7C and D). Because the localisation of U3 and U4 remained unaltered, the differential localisation of NHPX at different time points is likely not due to the movement or relocalisation of either

of these NHPX target RNAs. These data are consistent with NHPX binding U4 snRNA in speckles and U3 snoRNA in the nucleolus. However, I cannot exclude that NHPX also binds to different and/or unknown target RNAs at the different nuclear structures.

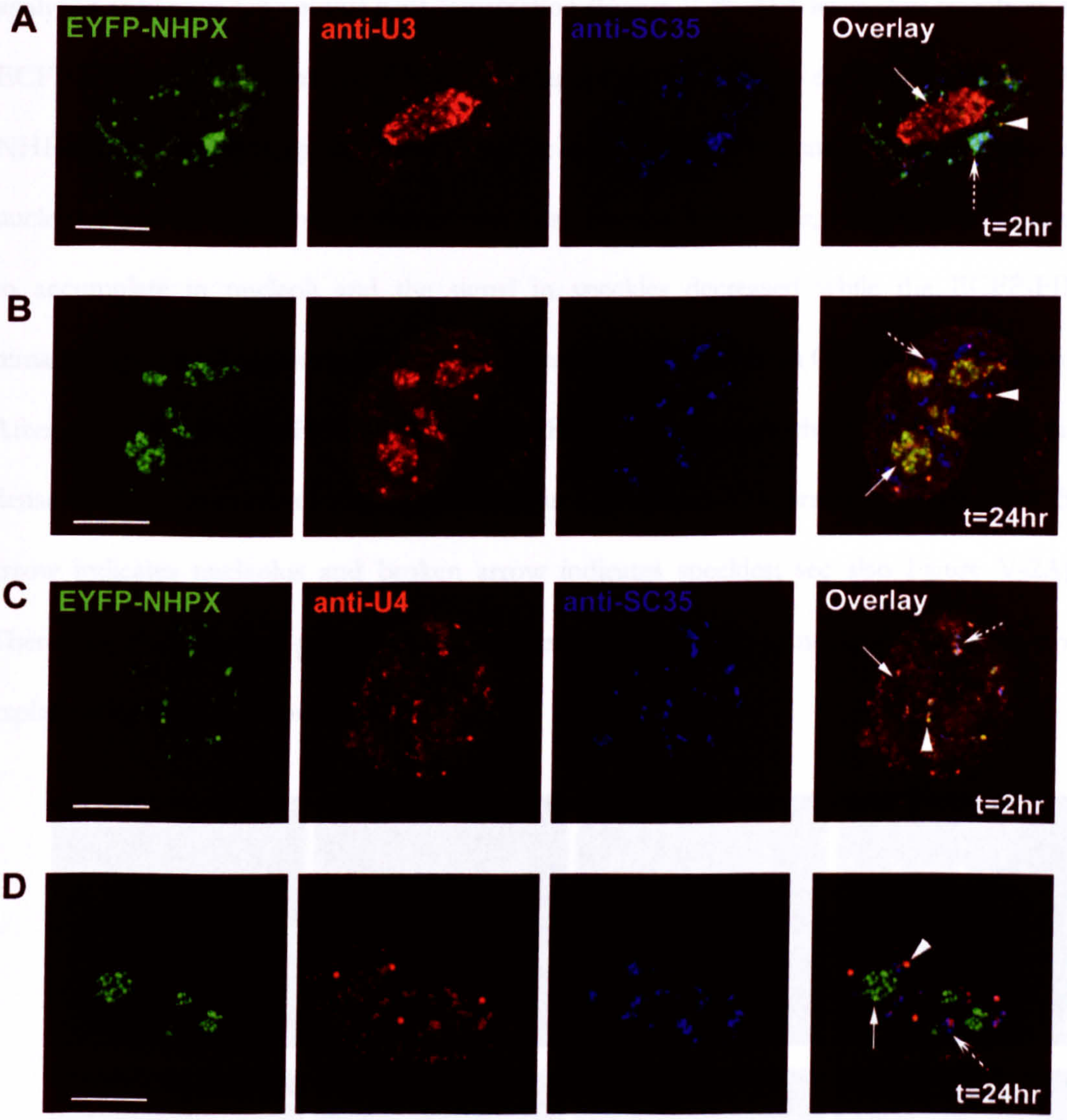


Figure V-7 **Endogenous RNA localisation in cells microinjected with plasmid expressing EYFP-NHPX.** pAL107$^{EYFP\text{-}NHPX}$ was microinjected into HeLa cells and fixed after (A, C) 2 and (B, D) 24 hours. The cells were then counterstained with antisense 2'-*O*-Methyl RNA (A, B) U3 and (C, D) U4 and anti-SC35 to show the location of speckles. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles. Scale bar=5μm

V.3.2 Evidence from co-transfections of plasmids expressing EYFP-NHPX and ECFP-FIB

I next examined whether the speckle localisation of newly synthesised NHPX is a consequence of a previously unknown behavior of U3 snoRNP. To test this, an expression vector, pAL118$^{ECFP\text{-}FIB}$, encoding ECFP-tagged fibrillarin (FIB), a known U3 snoRNP component, was cotransfected with pAL107$^{EYFP\text{-}NHPX}$ into the parental HeLa cell line and analyzed at various time points post-transfection (Figure V-8). At 1 hour after transfection, ECFP-FIB already accumulated in nucleoli and CBs, whereas in the same cells EYFP-NHPX accumulated in speckles and CBs, but not in nucleoli (Figure V-8; arrow indicates nucleolus and broken arrow indicates speckles). Several hours later, EYFP-NHPX began to accumulate in nucleoli and the signal in speckles decreased while the ECFP-FIB remained specifically in nucleoli and both proteins were detected in CBs (data not shown). After one cell cycle, EYFP-NHPX and ECFP-FIB both quantitatively colocalised in the dense fibrillar component inside nucleoli and in CBs (Figure V-8; arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles; see also Figure V-2A). Therefore, the transient presence of newly expressed NHPX in nuclear speckles is not explained by its association with U3 snoRNP.

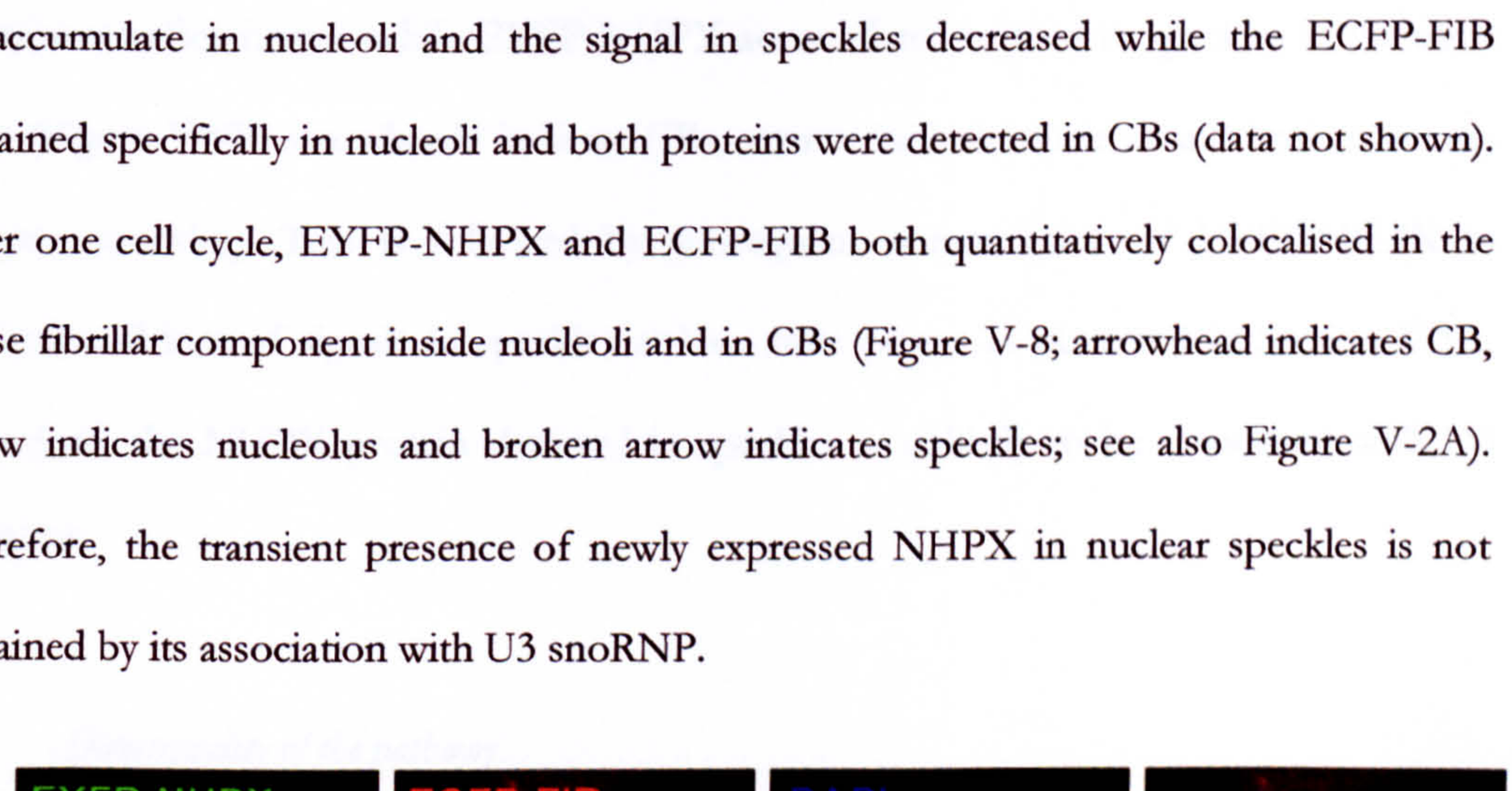


Figure V-8 **pAL107$^{EYFP\text{-}NHPX}$ and pAL118$^{ECFP\text{-}FIB}$ were co-transfected into HeLa cells and fixed after (A) 1 and (B) 24 hours.** Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles; Scale bar=5μm.

V.3.3 Evidence from heterokaryon assays between HeLa stable cell lines

The differential timing in the nucleolar entry of NHPX and fibrillarin is confirmed by analysis of heterokaryons formed between $HeLa^{ECFP\text{-}FIB}$ and $HeLa^{EYFP\text{-}NHPX}$ stable cell lines (Figure V-9). Formation of heterokaryons between two cell lines expressing FP-tagged proteins allows the gradual introduction of EYFP-NHPX into the $HeLa^{ECFP\text{-}FIB}$ cells and *vice versa* (Sleeman et al., 2001). The advantages of using this heterokaryon approach are that it minimizes the possible effect of overexpression that can occur in both microinjection and transient transfection assays and allows the dynamic exchange of two fluorophores to be observed under the same conditions. ECFP-FIB accumulated directly in nucleoli as soon as 15 minutes after fusion, while EYFP-NHPX accumulated instead in speckles at the same time (Figure V-9B; arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles). This is followed by subsequent accumulation of EYFP-NHPX in nucleoli and loss of signals in speckles at later time points as observed in Section V.3.1. Therefore, the NHPX protein detected in speckles is unlikely to be associated with U3 snoRNP.

V.4 Directionality of the pathway

V.4.1 Reciprocal movement of two different nuclear proteins

The NHPX pathway appears complementary to that previously described for Sm proteins (Figure V-10; Sleeman and Lamond, 1999; Sleeman et al., 2001). By making heterokaryons between $HeLa^{ECFP\text{-}SmB}$ and $HeLa^{EYFP\text{-}NHPX}$, I observed that EYFP-NHPX entered into speckles directly, whereas ECFP-SmB accumulated specifically in CBs shortly after fusion (Figure V-10A and B; arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles). At 2 hours after fusion, ECFP-SmB remained in CBs while the EYFP-NHPX signal inside nucleoli increased (Figure V-10C; arrows indicate nucleoli). In some cells, ECFP-SmB was also detected inside nucleoli, as reported previously (Figure V-10D,

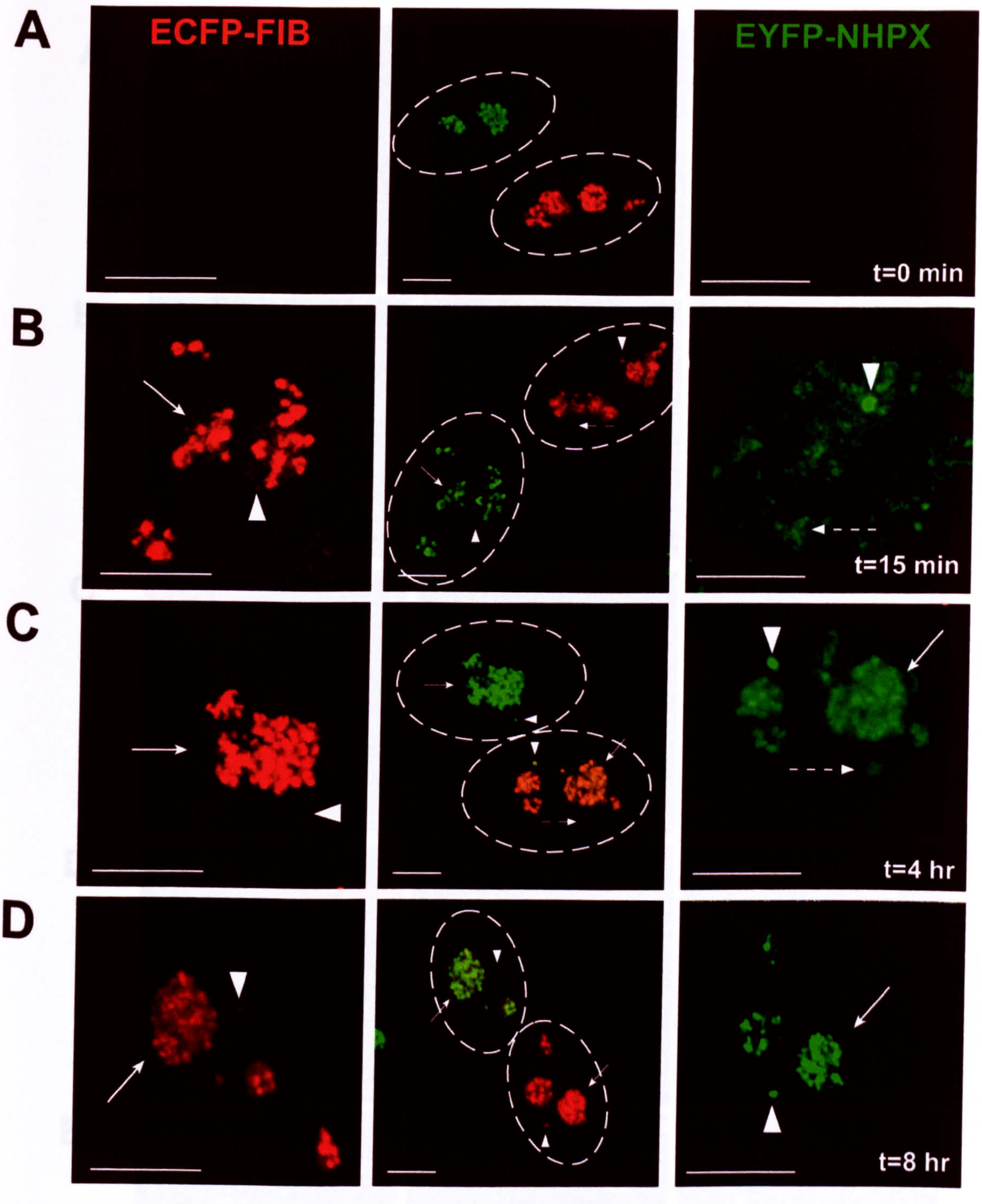


Figure V-9 **Heterokaryon assays between HeLa$^{EYFP\text{-}NHPX}$ and HeLa$^{ECFP\text{-}FIB}$.** HeLa$^{EYFP\text{-}NHPX}$ and HeLa$^{ECFP\text{-}FIB}$ were fused to form heterokaryon using polyethylene glycol (PEG) and were fixed at 0, 15, 240 and 480 minutes after fusion. In order to show the relative distribution of EYFP and ECFP components in each nucleus, the central panel shows the same two nuclei within a heterokaryon as the side panels, but in the opposite fluorescence channels. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles; dotted ovals outline nuclei in the heterokaryon in the central panel. Scale bar=5μm.

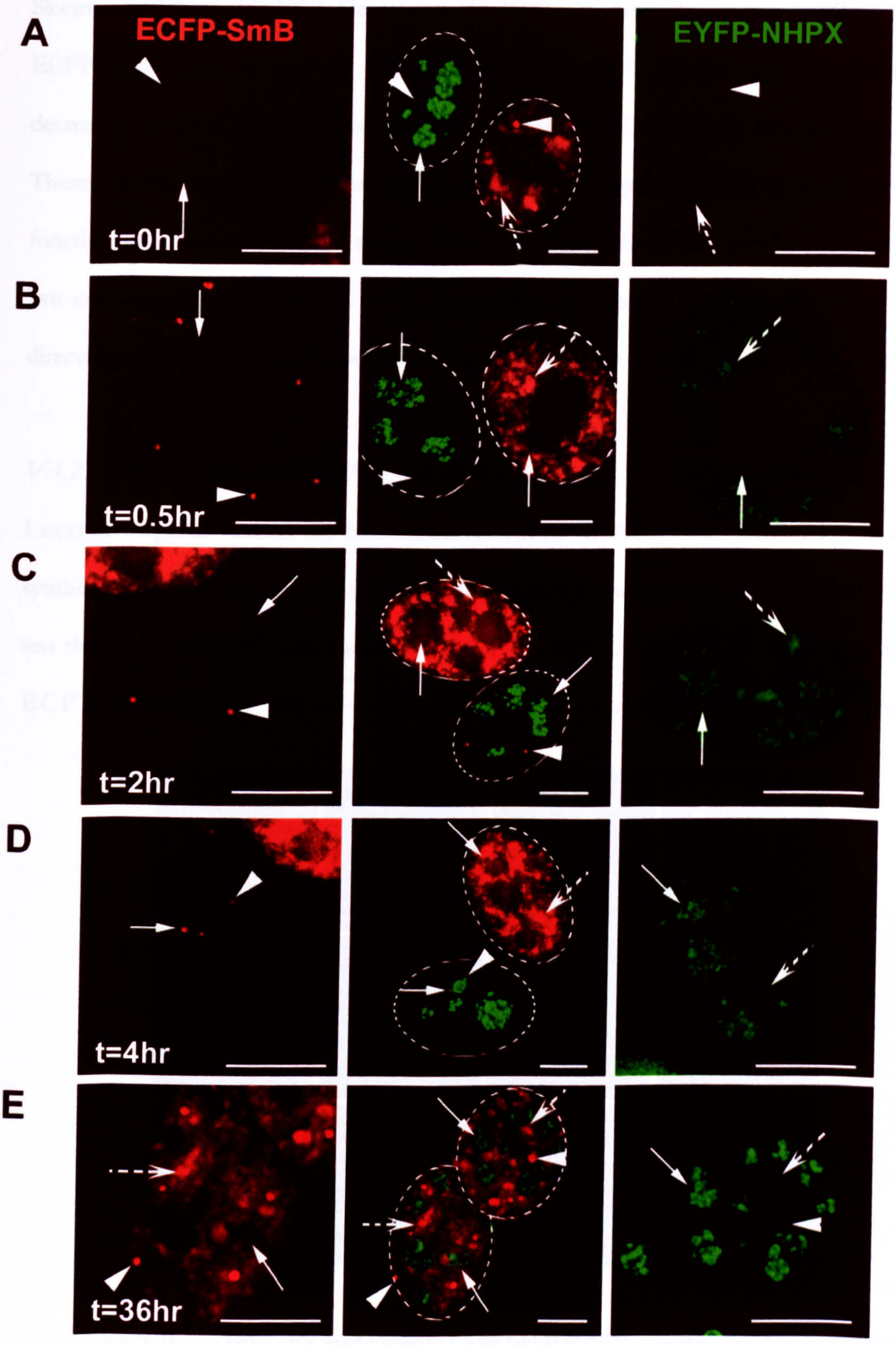


Figure V-10 **Reciprocal movement of nuclear proteins NHPX and SmB.** HeLa$^{EYFP\text{-}NHPX}$ and HeLa$^{ECFP\text{-}SmB}$ were fused to form heterokaryon using PEG and were fixed at different time points: (A) 0hr, (B) 0.5hr, (C) 2hr (D) 4hr and (E) 36hr. Panel representation as of Figure V-9. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles; dotted ovals outline nuclei of the heterokaryon in the central panel. Scale bar=5μm.

Sleeman and Lamond, 1999; Sleeman et al., 2001). At later time points (~36hrs), the ECFP-SmB signal in speckles increased while the EYFP-NHPX signal in speckles decreased to a very low/undetectable level (Figure V-10E and other data not shown). Therefore, I conclude that both nuclear pathways, though operating in different directions, function simultaneously inside a single cell nucleus. The pathways can also be observed by live cell imaging over a period of 12 hours (data not shown). This demonstrates the directed movement of proteins between separate, membrane free nuclear compartments.

V.4.2 Only newly expressed NHPX proteins localise to speckles

I next investigated whether the directed movement of NHPX is either restricted to newly synthesised proteins, or else is a reversible relocation of existing proteins (Figure V-11). To test this, pAL214$^{ECFP\text{-}NHPX}$ was transfected into HeLa cells and left for 24 hours such that ECFP-NHPX was already accumulated in nucleoli and CBs before

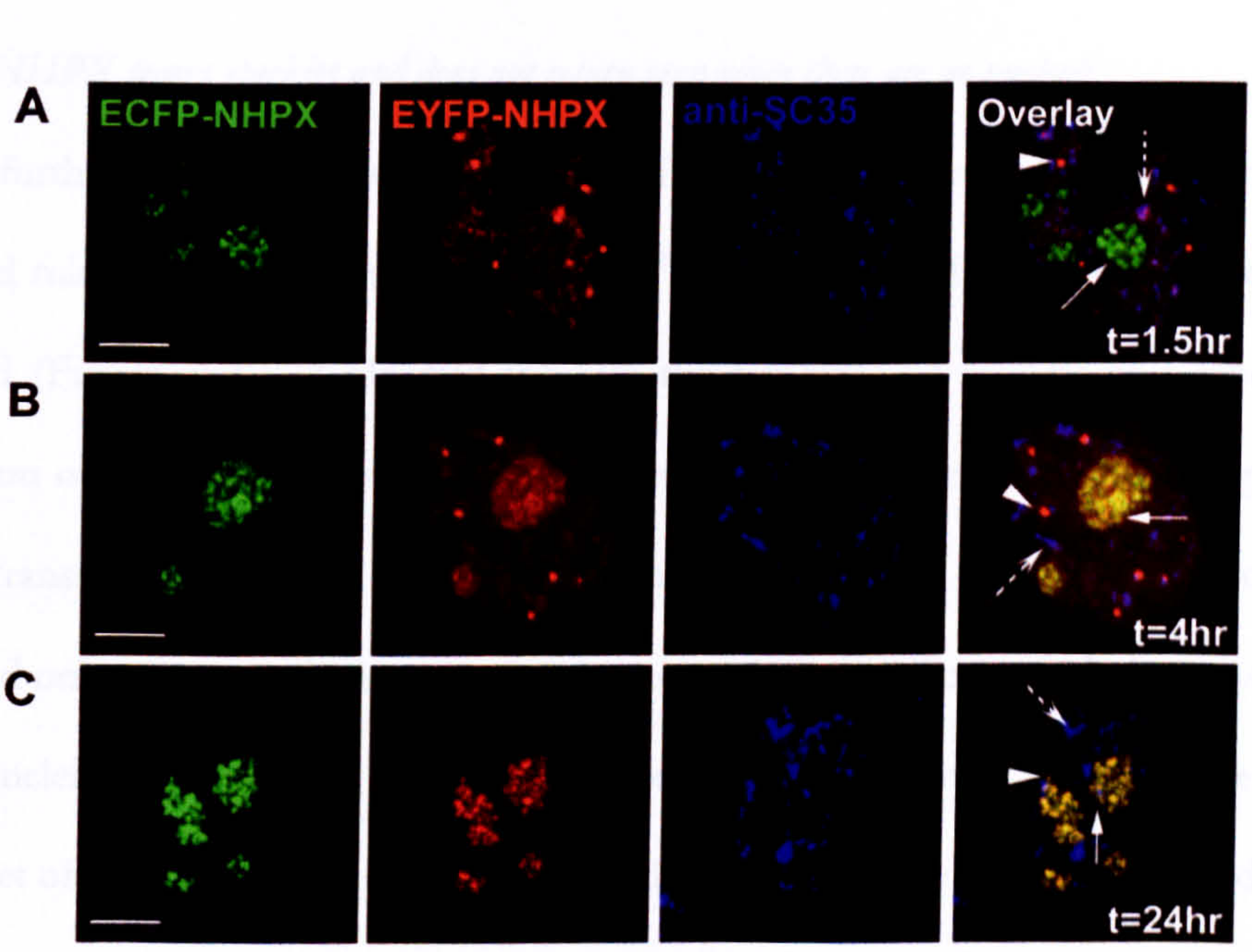


Figure V-11 **HeLa cells were transfected with pAL214$^{ECFP\text{-}NHPX}$ for 24hrs before microinjecting pAL107$^{EYFP\text{-}NHPX}$.** The microinjected cells were fixed after (A) 1.5hr, (B) 4hr and (C) 24hr and counterstained with anti-SC35 to denote speckles. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles. Scale bar=5μm

microinjection of pAL107$^{EYFP\text{-}NHPX}$. Microinjection of pAL107$^{EYFP\text{-}NHPX}$ provided a pulse of newly synthesised EYFP-NHPX that accumulated in splicing speckles and CBs while the previously expressed ECFP-NHPX accumulated in nucleoli and CBs instead (Figure V-11A; arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles). Gradually, EYFP-NHPX appeared in nucleoli, while the signal in speckles subsided. The nucleolar pattern of ECFP-NHPX remained unaltered after microinjection of pAL107$^{EYFP\text{-}NHPX}$ (Figure V-11B; arrow indicates nucleolus). At 24 hours post-microinjection, EYFP-NHPX completely colocalised with the existing ECFP-NHPX (Figure V-11C; arrowhead indicates CB and arrow indicates nucleolus). These data indicate that only newly synthesised NHPX accumulates in splicing speckles and further argue that this association is transient. The presence of NHPX in speckles is thus likely not a result of protein relocation due to exogenous expression.

V.4.3 NHPX leaves speckles and does not return even when there are no nucleoli

To test further whether or not pools of NHPX in speckles and nucleoli interchange, I generated micronuclei by treating the HeLa$^{EYFP\text{-}NHPX}$ cells with the spindle-disrupting drug colcemid (Figure V-12). Colcemid inhibits the progress of mitosis and renders the segregation of chromosomes into many micronuclei without preventing DNA replication, mRNA transcription, splicing and protein translation (Ferreira et al., 1997). Nucleoli are assembled only on the nucleolar organizer regions (NORs) in 5 out of 23 chromosomes in human nuclei (see Chapter I) and therefore the colcemid treatment allows the generation of a subset of micronuclei without nucleoli. To locate those micronuclei, I screened with an antibody specific for nucleolar antigen B23. If the two pools of NHPX in splicing speckles and nucleoli freely exchange, NHPX originally from nucleoli would be expected by default to accumulate back in splicing speckles in the micronuclei lacking nucleoli. Interestingly, EYFP-NHPX does not accumulate in splicing speckles, even in those micronuclei lacking

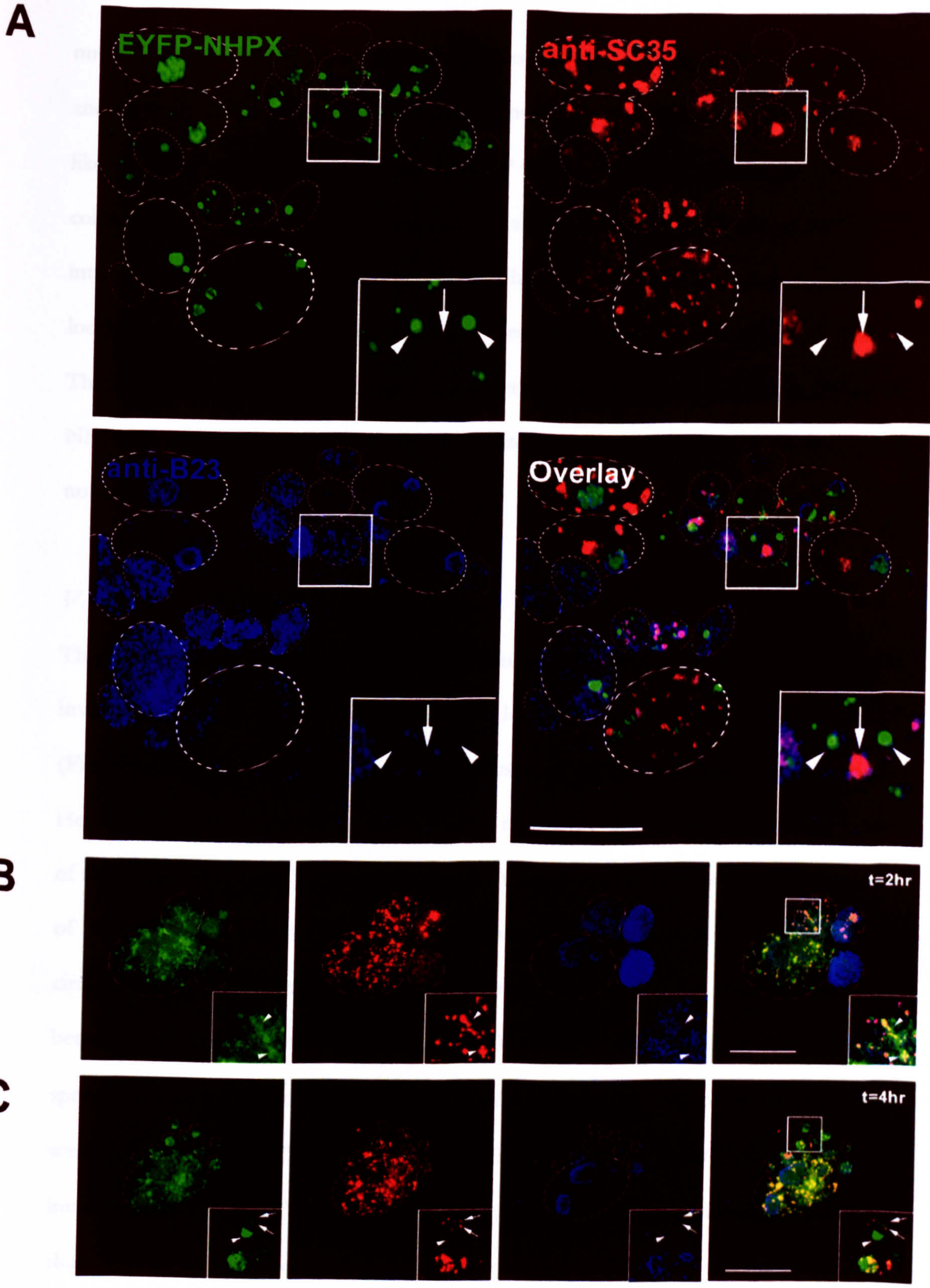


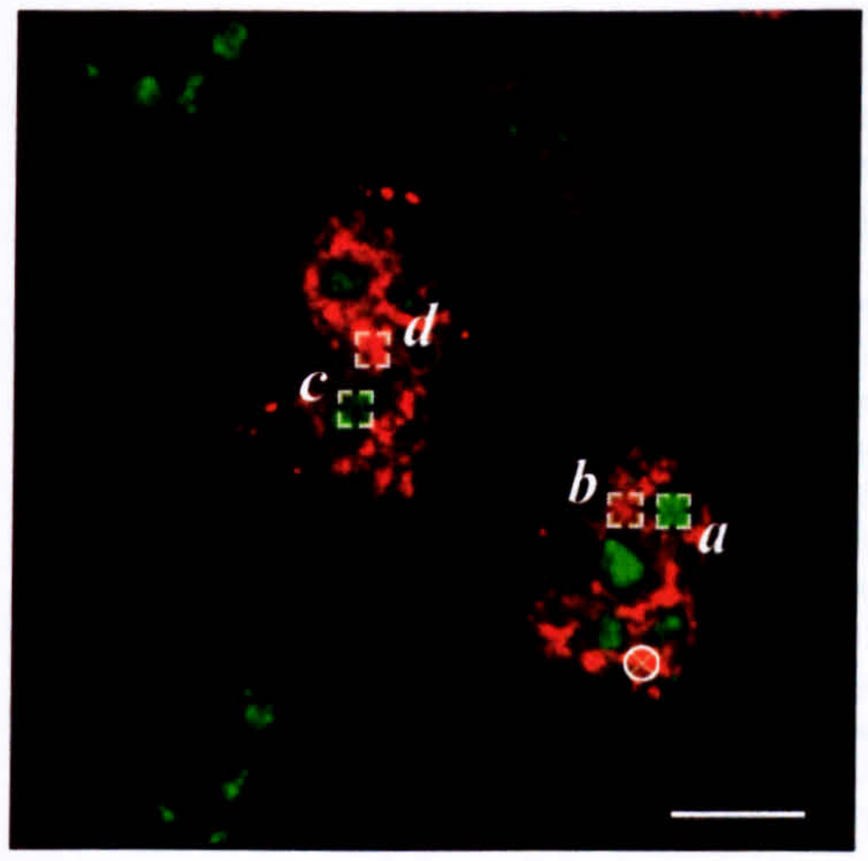
Figure V-12 **Localisation of NHPX in cells lacking nucleoli.** (A) HeLa$^{EYFP\text{-}NHPX}$ cells were fixed after treating with colcemid for 31hrs and counterstained with anti-SC35 to denote speckles and anti-B23 to denote nucleoli. (B-C) Parental HeLa cells were treated with colcemid for 31hr and microinjected with plasmid expressing EYFP-NHPX. The cells were fixed at (B) 2 hr and (C) 4 hr after microinjection. Arrows indicate speckles while arrowheads indicate the locations of EYFP-NHPX; dotted ovals outline the micronuclei. Scale bar=5μm.

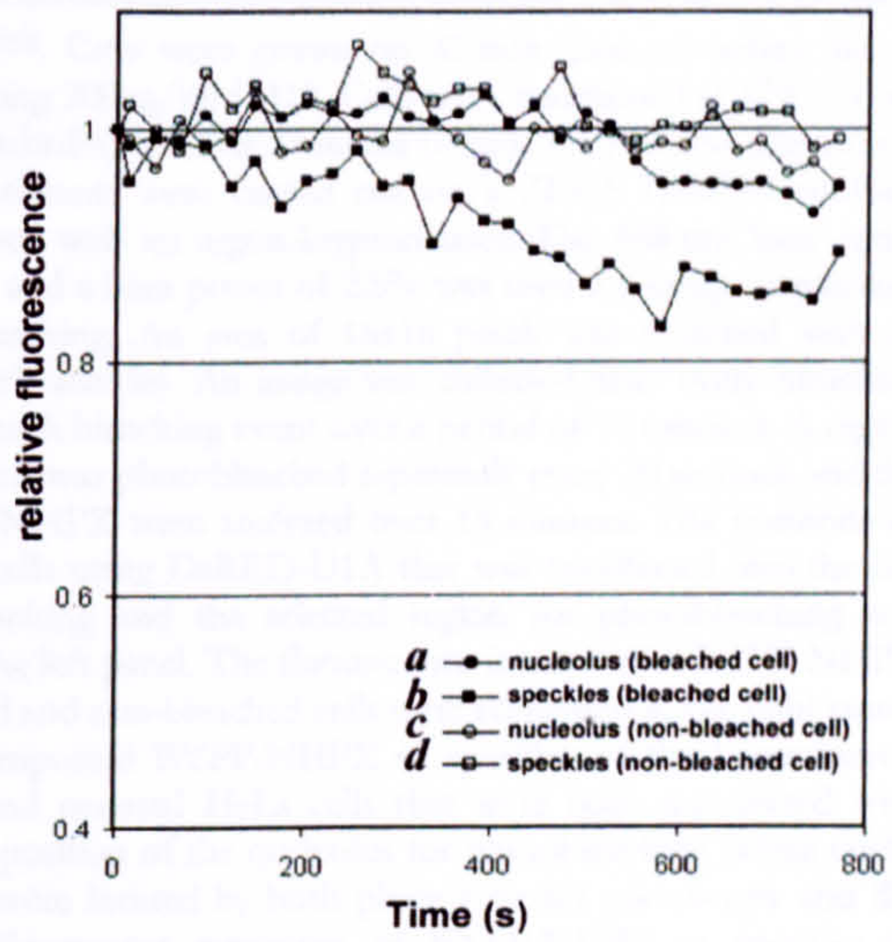
nucleoli (Figure V-12A; arrows indicate speckles, arrowheads indicate NHPX localisations and the inset shows a micronucleus that lacks nucleoli); instead, they are localised in spot-like structures which also contain the snoRNP protein fibrillarin, but not the CB marker coilin (Figure V-12A, arrowheads and data not shown). Microinjection of pAL107$^{EYFP\text{-}NHPX}$ into colcemid-treated parental HeLa cells showed the same temporal sequence of localisation in speckles prior to nucleoli as seen for untreated cells (Figure V-12B-C). Therefore, this differential localisation is apparently not a result of colcemid modifying the NHPX pathway and the pools of NHPX localised in splicing speckles and nucleoli appear not to interchange.

V.4.4 The NHPX pathway is unidirectional

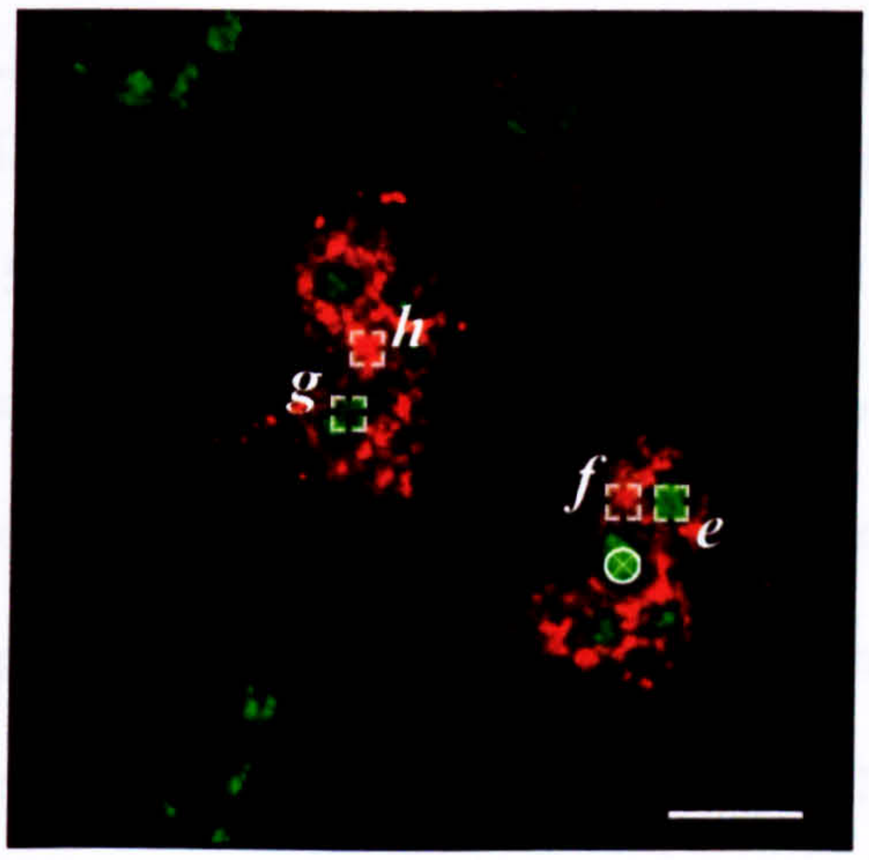
The non-cycling behavior of NHPX between speckles and nucleoli prompted me to investigate further the directionality of the localisation pathway. I performed FLIP (Fluorescence Loss In Photobleaching) analyses of different nuclear structures in the HeLa$^{EYFP\text{-}NHPX}$ cells where one area of the cell is repeatedly bleached while collecting images of the entire cell (Figure V-13; see Chapter I). If fluorescent molecules from other regions of the cell diffuse into the bleached area (Figure V-13, bleach zone is shown by white circle), loss of fluorescence will occur from both places, indicating that molecules exchange between the regions (Reits and Neefjes, 2001). First, I tested whether NHPX inside speckles is moving into the nucleolus (Figure V-13A; ***a*** and ***c***). The positions of speckles were defined by DsRED-U1A in live cells (Figure V-13A; ***b*** and ***d***). The fluorescence intensity of EYFP-NHPX in speckles outside the bleached region decreased, indicating that NHPX diffuses between these nuclear domains (Figure V-13A; curve ***b***). In comparison, the signals inside nucleoli only showed a minor decrease (Figure V-13A; curve ***a***). This is consistent with the expected movement of NHPX from speckles to nucleoli. The small change in nucleolar fluorescence may be because the directed movement of

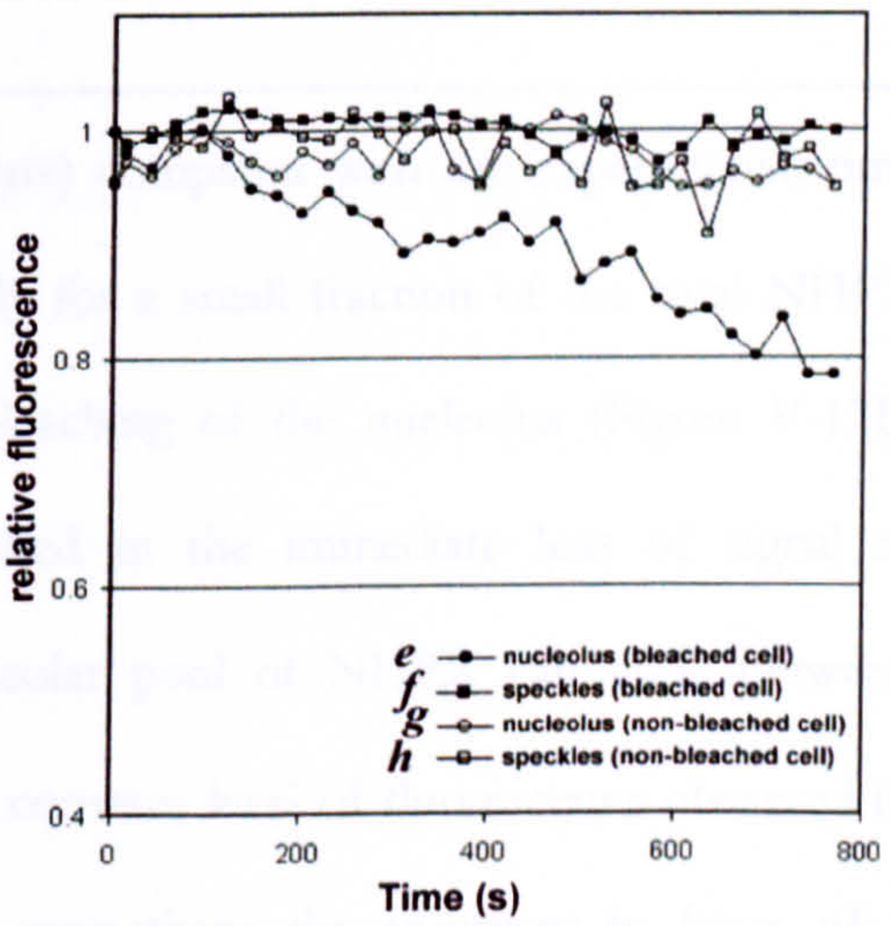
A


d
c
b
a

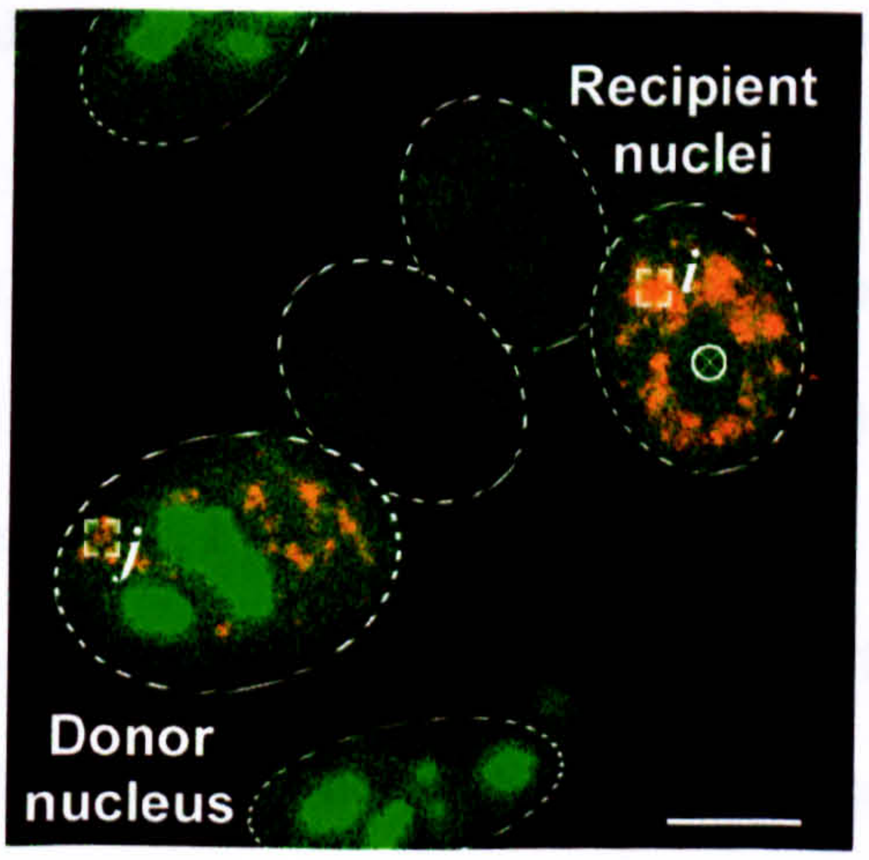

relative fluorescence
1
0.8
0.6
0.4
0
200
400
600
800
Time (s)
a nucleolus (bleached cell)
b speckles (bleached cell)
c nucleolus (non-bleached cell)
d speckles (non-bleached cell)

B


h
g
f
e


relative fluorescence
1
0.8
0.6
0.4
0
200
400
600
800
Time (s)
e nucleolus (bleached cell)
f speckles (bleached cell)
g nucleolus (non-bleached cell)
h speckles (non-bleached cell)

C


Recipient nuclei
i
j
Donor nucleus

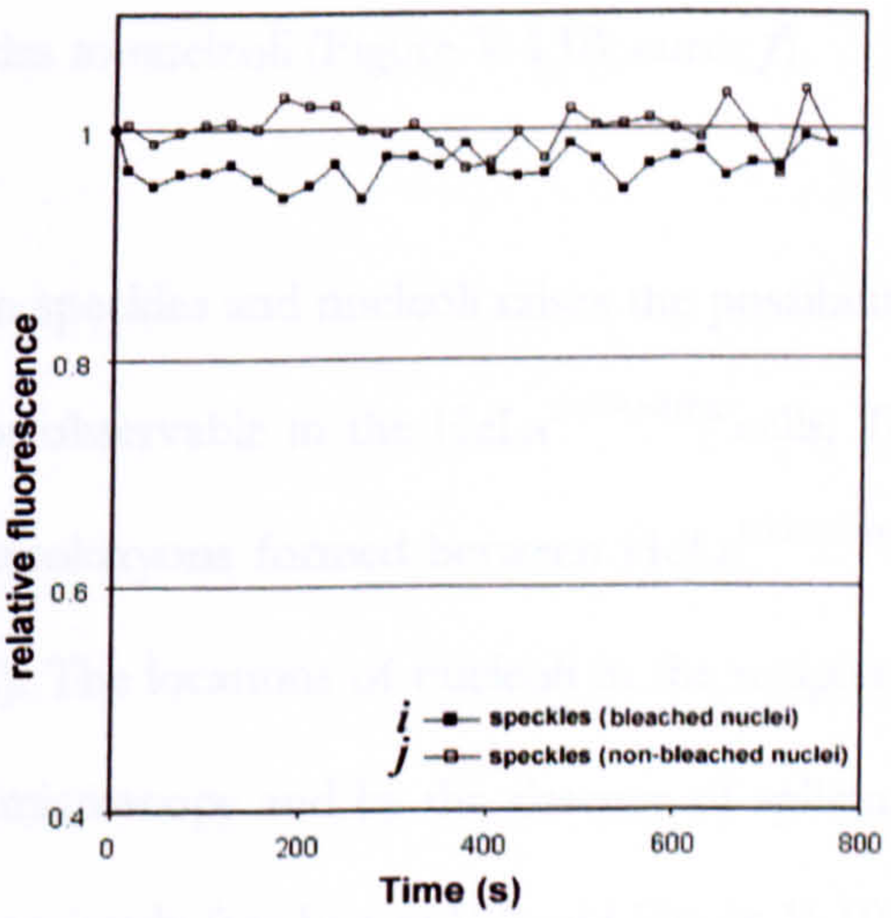

relative fluorescence
1
0.8
0.6
0.4
0
200
400
600
800
Time (s)
i speckles (bleached nuclei)
j speckles (non-bleached nuclei)

Figure V-13 **FLIP analysis in HeLa$^{EYFP\text{-}NHPX}$.** Cells were grown on 42-mm glass coverslips (no. 1; Helmut Sauer) in medium containing 200μg/ml G418. Cells were maintained at 37°C by use of a closed perfusion chamber (Bachofer) in DMEM media (20mM HEPES, no phenol red; Invitrogen). Photobleaching experiments were carried out on a ZEISS LSM510 confocal laser scanning microscope equipped with an argon-krypton laser. The 488-nm laser and a 63X plan Apo lens with a 1.4 NA and a laser power of 2.5% was used for image acquisition, and 25% was used for photobleaching. An area of 16x16 pixels was bleached with an iteration of 250 (duration of bleach was 3s). An image was collected after every bleaching event, with 20s intervals between each bleaching event over a period of 15 minutes. A region in the (A) speckles and (B) nucleolus was photobleached repeatedly every 20 seconds and the fluorescence intensities of EYFP-NHPX were analyzed over 15 minutes. The positions of speckles were located in the live cells using DsRED-U1A that was transfected into the cell lines for 24hrs before photobleaching and the selected region for photobleaching was highlighted by the white circle in the left panel. The fluorescence intensities of EYFP-NHPX in different regions of the bleached and non-bleached cells were compared in the right panel. (C) FLIP analysis of the newly imported EYFP-NHPX in speckles of the heterokaryon formed between Hela$^{EYFP\text{-}NHPX}$ and parental HeLa cells that were both transfected with pDsRED-U1A for 24 hours. The position of the nucleolus for photobleaching (white circle; left panel) in the recipient nuclei were located by both phase contrast microscopy and the absence of DsRED-U1A. The fluorescent intensities of EYFP-NHPX in speckles of bleached and non-bleached nuclei of the heterokaryon were analyzed and shown on the right panel. Dotted ovals outline nuclei in the heterokaryon. Scale bar=5μm.

NHPX from speckles to nucleoli is slow (hours) compared with the experimental time (~15 minutes) and/or because it accounts only for a small fraction of the total NHPX signal in nucleoli. However, repeated photobleaching of the nucleolus (Figure V-13B; bleach zone is shown by white circle) resulted in the immediate loss of signal in neighboring nucleoli, indicating that the nucleolar pool of NHPX can cycle between different nucleoli (Figure V-13B; curve *e*). The constant level of fluorescence observed in the speckles in the same experiment further strengthens the argument in favor of a unidirectional movement of NHPX from speckles to nucleoli (Figure V-13B; curve *f*).

The difference in fluorescence intensity between speckles and nucleoli raises the possibility that the flow from nucleoli to speckles was not observable in the HeLa$^{EYFP\text{-}NHPX}$ cells. To address this, I performed FLIP analysis on heterokaryons formed between HeLa$^{EYFP\text{-}NHPX}$ cells and the parental HeLa cells (Figure V-13C). The locations of nucleoli in the recipient nuclei were positioned both by phase contrast microscopy and by the absence of splicing factor U1A. Shortly after the fusion, as shown previously (see Figure V-9 and Figure V-10), EYFP-NHPX first appeared in splicing speckles but was absent from nucleoli in the

recipient nuclei. Repeated photobleaching of the nucleoli did not change the fluorescence level of EYFP-NHPX in the speckles of the recipient nuclei (Figure V-13C; curve ***i*** and bleach zone is shown by white circle). Therefore, NHPX either does not cycle from nucleoli to speckles, or else does so at a rate/level that cannot be detected in this assay. In summary, the photobleaching analyses, combined with the other data presented here, suggest a unidirectional movement of NHPX from splicing speckles to nucleoli. However, pools of EYFP-NHPX appear freely diffusible between separate components of the same nuclear structure, indicating the regulated entry of nuclear proteins into different domains inside the nucleus.

V.4.5 The progression of NHPX from speckles to nucleoli is dependent on RNA polymerase II but not RNA polymerase I transcription

I next tested whether the progression of NHPX from speckles to nucleoli requires gene expression, including both RNA polymerase I and RNA polymerase II transcription (Figure V-14). I again employed the heterokaryon approach between $HeLa^{ECFP\text{-}FIB}$ and $HeLa^{EYFP\text{-}NHPX}$ cells. At 4 hour post-fusion, only a small amount of EYFP-NHPX was in speckles whilst most accumulated in nucleoli (Figure V-14A; arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles; see also Figure V-10D). The heterokaryons formed between $HeLa^{EYFP\text{-}NHPX}$ and $HeLa^{ECFP\text{-}FIB}$ cells were subjected to transcription inhibitors targeted to specific polymerases. Low levels of actinomycin D cause the segregation of nucleoli and inhibit rRNA transcription, but not pre-mRNA transcription (see Chapter I). Newly synthesised EYFP-NHPX moved to the speckles of recipient $HeLa^{ECFP\text{-}FIB}$ nuclei, prior to accumulating in the segregated nucleoli (Figure V-14B; arrows indicate segregated nucleoli and broken arrows indicate speckles), suggesting that RNA polymerase I transcription and/or ribosome biogenesis is not a prerequisite for the NHPX pathway. Similarly, the immunosuppressant rapamycin, which

inhibits transcription of a subset of ribosomal protein genes and hence ribosome assembly, gave the same results (data not shown). However, when RNA polymerase II transcription was inhibited, either by α-amanitin or DRB, progression of the newly synthesised NHPX from speckles to the nucleolus was blocked (Figure V-14C and data not shown; arrows indicate nucleoli and broken arrows indicate speckles). This suggests that one or more factors must be continually synthesised to allow the newly imported NHPX to move from speckles to nucleoli.

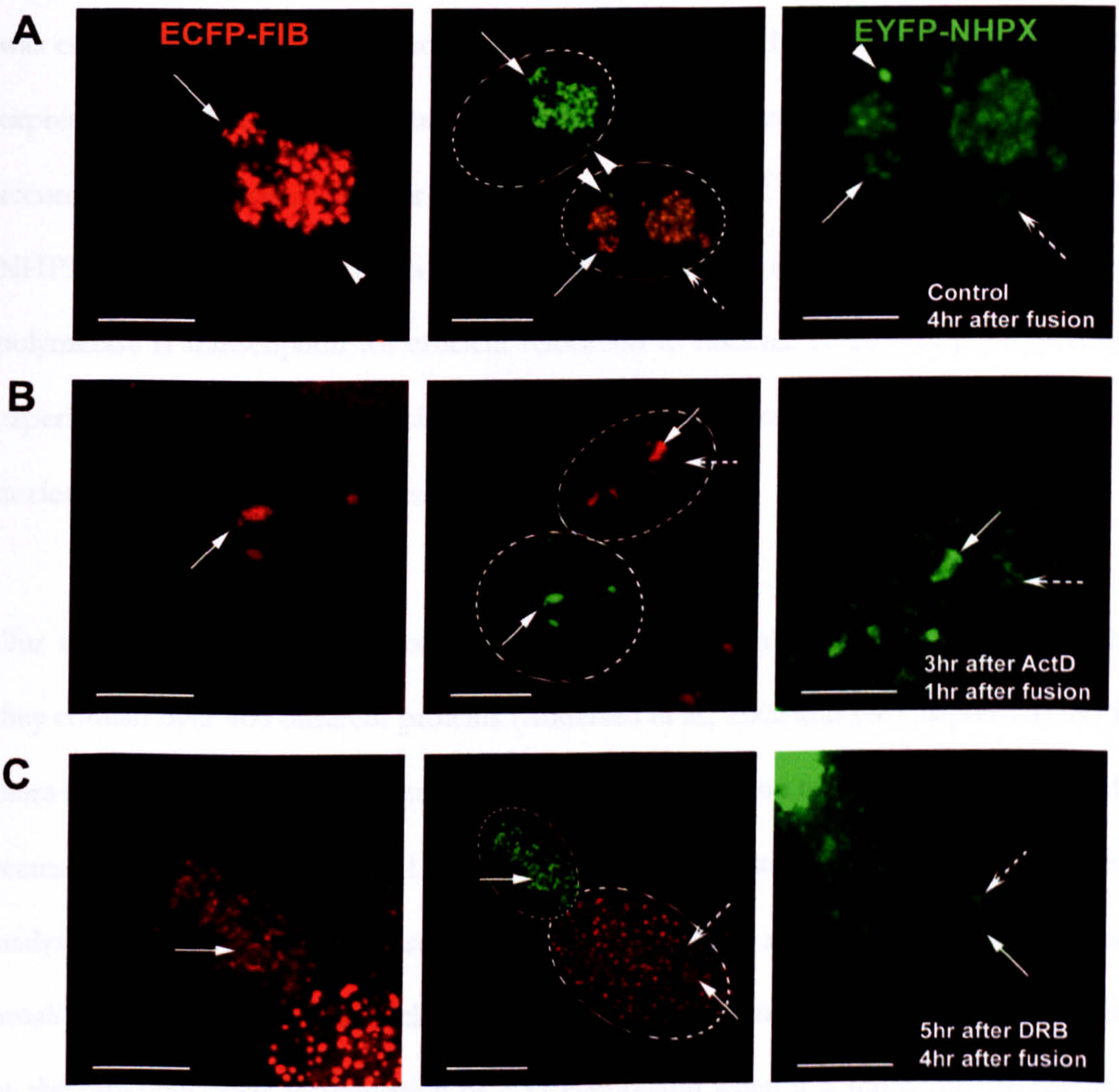


Figure V-14 **The NHPX pathway is dependent on RNA polymerase II transcription.** Heterokaryons formed between HeLa$^{EYFP\text{-}NHPX}$ and HeLa$^{ECFP\text{-}FIB}$ were treated with different transcription inhibitors: (A) control, (B) RNA polymerase I inhibitor Actinomycin D (0.04μg/ml) and (C) RNA polymerase II inhibitor DRB (100μM). Panel representation as of Figure V-9. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles; dotted ovals outline nuclei of the heterokaryon in the central panel. Scale bar=5μm.

V.5 Discussion

In this chapter, I described a novel nuclear pathway that I discovered and published during my PhD studies (Leung and Lamond, 2002 and see Appendix). The pathway that leads to the nucleolar accumulation of the NHPX protein was detected in multiple mammalian cultured cell lines, including both primary and transformed cells. NHPX was analyzed *in vivo*, fused to either EYFP or ECFP fluorescent protein tags, and the resulting fusion proteins were shown to have similar localisation patterns and RNA binding specificities to the endogenous NHPX. A stably transformed HeLa cell line that expressed EYFP-NHPX was established and used to demonstrate that, upon its initial entry into the nucleus, newly expressed NHPX transiently accumulates in splicing speckles prior to a later, steady state accumulation in nucleoli. Further characterisation of HeLa$^{EYFP\text{-}NHPX}$ cells indicated that the NHPX protein in speckles was not associated with U3 snoRNP and required RNA polymerase II transcription for efficient relocation to nucleoli. Additional photobleaching experiments showed that the nucleolar pool of NHPX did not interchange with the pool in nuclear speckles, suggesting a unidirectional pathway.

Our recent proteomic analysis of nucleoli isolated from cultured HeLa cells shows that they contain over 400 different proteins (Andersen et al., 2002 and see Chapter III). So far there has been no nucleolar targeting motif identified common to all of these factors and it seems likely that multiple, parallel nucleolar localisation pathways can operate. Nonetheless, analyses of proteins that show a steady state accumulation in nucleoli have shown that they usually move rapidly into the nucleolus when they enter the nucleus. This is illustrated here by the rapid nucleolar accumulation of EYFP-fibrillarin when it is transiently expressed *in vivo* (cf. Figure V-8 & Figure V-9). The finding that newly expressed NHPX accumulates in nuclear speckles transiently before accumulating specifically in the nucleolus defines a new localisation pathway for nucleolar proteins. It is interesting to compare this with the

recently reported pathway for nucleolar localisation of snoRNAs, which showed that multiple snoRNAs accumulate in CBs prior to nucleoli upon initial entry into the nucleus (Narayanan et al., 1999a; Narayanan et al., 1999b; Verheggen et al., 2001). None of the snoRNAs, however, showed a transient accumulation in speckles, consistent with our finding that the snoRNP protein fibrillarin also does not transiently accumulate in speckles prior to nucleoli (cf. Figure V-8 & Figure V-9). I also observed that NHPX localises to CBs as well as speckles upon its initial entry into the nucleus, but unlike its transient association with speckles, NHPX is also detected in CBs at later stages of expression, when the bulk of the protein is concentrated in nucleoli. At present, I cannot distinguish whether NHPX either accumulates in CBs prior to speckles, or in both structures at the same time. However, the CB association does not appear to be obligatory for the NHPX localisation pathway because a similar transient association with speckles prior to nucleolar accumulation is observed in cell lines lacking prominent CBs (cf. Figure V-6). Yet, similar molecular events may occur, either within the nucleoplasm, or in CBs that are too small to detect.

It is also interesting to compare the NHPX pathway with the recently identified nuclear pathway for splicing snRNPs, where FP-tagged snRNP Sm proteins accumulate in CBs and nucleoli, prior to speckles, upon their initial nuclear entry (Sleeman and Lamond, 1999; Sleeman et al., 2001). This pathway therefore appears to be complementary to that of NHPX. By analyzing heterokaryons formed between separate stable HeLa cell lines expressing EYFP-NHPX and ECFP-SmB I could show that both these complementary pathways can operate simultaneously within the same nuclei (cf. Figure V-10). These data confirm the specificity of the pathways and highlight the dynamic mechanisms operating to organise the distribution of proteins and RNPs in the nucleus. The results also point to the

localisation specificity of the separate subnuclear bodies, including nucleoli, CBs and speckles, even although they are not enclosed by membranes.

In trying to answer why NHPX shows the observed transient accumulation in speckles prior to nucleoli, it may be important to consider that it is specifically newly expressed and imported NHPX protein that is detected in speckles. Several experiments showed that NHPX does not localise to speckles by default and that the nucleolar pool of NHPX does not cycle continually to and from speckles. For example, in micronuclei that lack NOR-containing chromosomes and hence do not have nucleoli, NHPX does not accumulate back in speckles or colocalise with splicing factors (cf. Figure V-12). FLIP photobleaching experiments also showed that while nucleolar EYFP-NHPX can exchange rapidly between separate nucleoli within the same nucleus, little or no exchange occurs with the pool of NHPX in speckles (cf. Figure V-13). This contrasts with the behavior of the nucleolar protein PSP1, which was recently shown to cycle continually between nucleoli and paraspeckles (Fox et al., 2002).

Our data strongly indicate that the association of NHPX with speckles is a temporal phenomenon linked to the entry of newly expressed NHPX into the nucleus. For example, transient expression of EYFP-NHPX in HeLa cells expressing ECFP-NHPX, which already accumulated in nucleoli, shows a transient accumulation of the EYFP-NHPX in speckles before it later colocalises quantitatively with the existing nucleolar ECFP-NHPX (cf. Figure V-11). Also, when the HeLa$^{EYFP\text{-}NHPX}$ cells undergo mitosis EYFP-NHPX immediately relocalises to the reforming nucleoli during telophase and does not accumulate in speckles in the post-mitotic nuclei (cf. Figure V-4). Therefore, the speckle association is not a result of nuclear import of NHPX *per se*, but rather relates to an effect specific for newly expressed protein. I propose that a likely explanation for this behavior of NHPX

could be related to it having a function required for the assembly or maturation of some form of nuclear protein or RNP complex, prior to its subsequent stable association with U3 and/or other nucleolar snoRNPs. This could imply either that the affinity of NHPX for different target RNAs changes after it enters the nucleus for the first time or that its access to bind snoRNA targets is initially restricted.

Based upon the results of previous biochemical studies on the structure and binding specificity of NHPX, the U4 snRNA is a possible candidate target for NHPX in speckles. NHPX binds U4 snRNA *in vitro* via the 5' stem loop sequence (Nottrott et al., 1999; Vidovic et al., 2000). Consistent with this idea, U4 snRNA has been localised to speckles in HeLa cells by hybridisation experiments (Carmo-Fonseca et al., 1992). The fact that I show here that EYFP-NHPX likely interacts *in vivo* with a form of U4 snRNA that is not stably associated with U6 snRNA suggests that NHPX may transiently interact in speckles with an intermediate form of U4 snRNP (cf. Figure V-2E).

In mammalian cells, all snoRNAs that guide 2'-O-methylation or pseudouridylation of ribosomal RNAs are encoded within the introns of snoRNA host genes (Maxwell and Fournier, 1995; Weinstein and Steitz, 1999; Filipowicz and Pogacic, 2002). They are processed exonucleolytically after debranching of the excised introns (Kiss and Filipowicz, 1993; Kiss and Filipowicz, 1995; Caffarelli et al., 1996) and the distance (>41nt) from the snoRNA coding region to the branch point is crucial for the release of a snoRNA from its host intron before the first catalytic step, as illustrated in Figure V-15 (Hirose and Steitz, 2001). It should be noted that NHPX is the nucleation factor for both the U4 splicing snRNP and also snoRNPs (Watkins et al., 2000; Nottrott et al., 2002; Watkins et al., 2002). During the splicing cycle, U4/U6.U5 tri-snRNP was recruited to the exon-intron junction to activate the spliceosome for the first catalytic step. When U2 is base-paired with U6

snRNA, the U4/U6 base pairing must first be disassociated and the fate of the U4 snRNP remains unknown. Recently, the Steitz group showed that the binding of fibrillarin and NHPX proteins to the snoRNA coding region occurs before the first catalytic step, and that snoRNP assembly *in vitro* is completely abolished by shortening the distance between the snoRNA coding region to the branch point to less than 60nt. (Hirose and Steitz 2001 and personal communication). I showed here that NHPX to be relocated from splicing speckles to nucleoli requires ongoing pre-mRNA transcription (c.f. Figure V-14). Therefore, it is tempting to speculate that NHPX molecules may be transferred from the U4 snRNP to snoRNPs at the first catalytic step when splicing snoRNA containing introns and this may explain the localisation pathway of NHPX involving a transient interaction with splicing speckles. Future studies will aim to analyze further the molecular mechanism involved in the novel nucleolar localisation pathway detected for NHPX *in vivo* and to establish what biological role this may play.

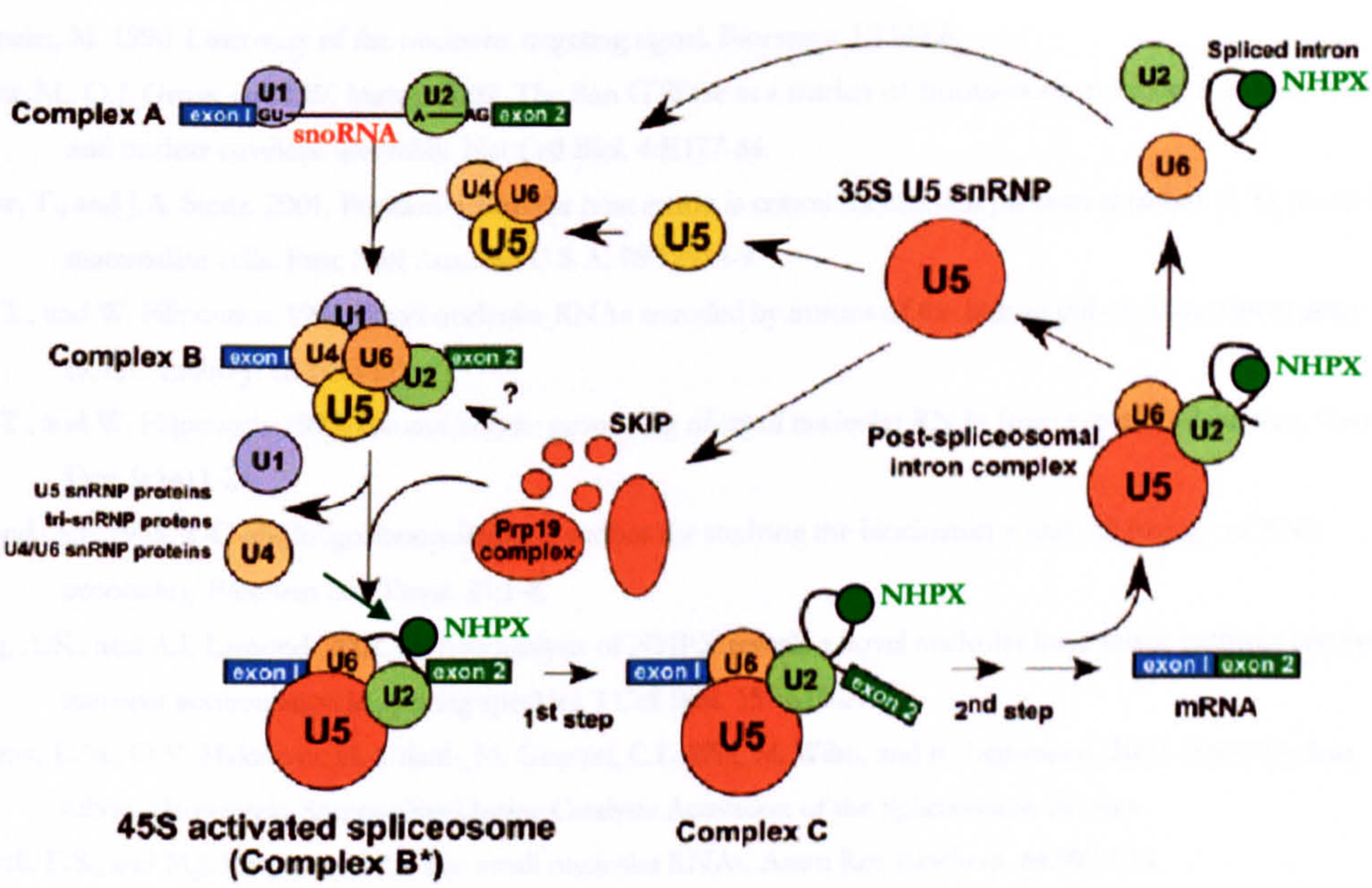


Figure V-15 **Splicing cycle.** The process of mRNA splicing begins with the binding of U1 and U2 at 5' and 3' exon-intron junction (Complex A), followed by the recruitment of tri-snRNP U4.U6/U5 (Complex B). The splicesomes is primed to be activated by the complementary base pairing between U2 and U6 snRNA at 3' end for the 1st and 2nd catalytic steps (Complex C), followed by the removal of intron lariats . It should be noted that the fate of the U4 snRNP at the stage when Complex B*/C formed remains unknown. NHPX from U4 snRNP free of U6 snRNA may be redistributed to intronic snoRNAs. The figure is adapted from Makarov et al., 2002.

V.6 References

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. Curr Biol. 12:1-11.

Boudonck, K., L. Dolan, and P.J. Shaw. 1999. The movement of coiled bodies visualised in living plant cells by the green fluorescent protein. Mol Biol Cell. 10:2297-307.

Burke, B., and J. Ellenberg. 2002. Remodelling the walls of the nucleus. Nat Rev Mol Cell Biol. 3:487-97.

Caffarelli, E., A. Fatica, S. Prislei, E. De Gregorio, P. Fragapane, and I. Bozzoni. 1996. Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. Embo J. 15:1121-31.

Carmo-Fonseca, M., R. Pepperkok, M. Carvalho, and A. Lamond. 1992. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. J Cell Biol. 117:1-14.

Carvalho, T., F. Almeida, A. Calapez, M. Lafarga, M.T. Berciano, and M. Carmo-Fonseca. 1999. The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. J Cell Biol. 147:715-28.

Chang, M.S., H. Sasaki, M.S. Campbell, S.K. Kraeft, R. Sutherland, C.Y. Yang, Y. Liu, D. Auclair, L. Hao, H. Sonoda, L.H. Ferland, and L.B. Chen. 1999. HRad17 colocalizes with NHP2L1 in the nucleolus and redistributes after UV irradiation. J Biol Chem. 274:36544-9.

Ferreira, J., G. Paolella, C. Ramos, and A.I. Lamond. 1997. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. J Cell Biol. 139:1597-610.

Filipowicz, W., and V. Pogacic. 2002. Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol. 14:319-27.

Fox, A.H., Y.W. Lam, A.K. Leung, C.E. Lyon, J. Andersen, M. Mann, and A.I. Lamond. 2002. Paraspeckles. A novel nuclear domain. Curr Biol. 12:13-25.

Hatanaka, M. 1990. Discovery of the nucleolar targeting signal. Bioessays. 12:143-8.

Hetzer, M., O.J. Gruss, and I.W. Mattaj. 2002. The Ran GTPase as a marker of chromosome position in spindle formation and nuclear envelope assembly. Nat Cell Biol. 4:E177-84.

Hirose, T., and J.A. Steitz. 2001. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. Proc Natl Acad Sci U S A. 98:12914-9.

Kiss, T., and W. Filipowicz. 1993. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. Embo J. 12:2913-20.

Kiss, T., and W. Filipowicz. 1995. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. Genes Dev. 9:1411-24.

Lamond, A.I. 1993. 2'-O-alkyloligoribonucleotides: probes for studying the biochemistry and cell biology of RNA processing. Biochem Soc Trans. 21:1-8.

Leung, A.K., and A.I. Lamond. 2002. In vivo analysis of NHPX reveals a novel nucleolar localization pathway involving a transient accumulation in splicing speckles. J Cell Biol. 157:615-29.

Makarov, E.M., O.V. Makarova, H. Urlaub, M. Gentzel, C.L. Will, M. Wilm, and R. Luhrmann. 2002. Small Nuclear Ribonucleoprotein Remodeling During Catalytic Activation of the Spliceosome. Science.

Maxwell, E.S., and M.J. Fournier. 1995. The small nucleolar RNAs. Annu Rev Biochem. 64:897-934.

Misteli, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. Science. 291:843-7.

Muratani, M., D. Gerlich, S.M. Janicki, M. Gebhard, R. Eils, and D.L. Spector. 2001. Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus. Nat Cell Biol. 21:21.

Narayanan, A., A. Lukowiak, B.E. Jady, F. Dragon, T. Kiss, R.M. Terns, and M.P. Terns. 1999a. Nucleolar localization signals of box H/ACA small nucleolar RNAs. Embo J. 18:5120-30.

Narayanan, A., W. Speckmann, R. Terns, and M.P. Terns. 1999b. Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. Mol Biol Cell. 10:2131-47.

Nottrott, S., K. Hartmuth, P. Fabrizio, H. Urlaub, I. Vidovic, R. Ficner, and R. Luhrmann. 1999. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. Embo J. 18:6119-33.

Nottrott, S., H. Urlaub, and R. Luhrmann. 2002. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. Embo J. 21:5527-38.

Okuwaki, M., A. Iwamatsu, M. Tsujimoto, and K. Nagata. 2001. Identification of nucleophosmin/B23, an acidic nucleolar protein, as a stimulatory factor for in vitro replication of adenovirus DNA complexed with viral basic core proteins. J Mol Biol. 311:41-55.

Phair, R.D., and T. Misteli. 2001. Kinetic modelling approaches to in vivo imaging. Nat Rev Mol Cell Biol. 2:898-907.

Platani, M., I. Goldberg, J.R. Swedlow, and A.I. Lamond. 2000. In vivo analysis of Cajal body movement, separation, and joining in live human cells. J Cell Biol. 151:1561-74.

Reddy, R., and H. Bush. 1988. Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. In Small Nuclear Ribonucleoprotein Particles. M.L. Bimstiel, editor. Springer-Verlag. 1-37.

Reits, E.A., and J.J. Neefjes. 2001. From fixed to FRAP: measuring protein mobility and activity in living cells. Nat Cell Biol. 3:E145-7.

Saito, H., T. Fujiwara, S. Shin, K. Okui, and Y. Nakamura. 1996. Cloning and mapping of a human novel cDNA (NHP2L1) that encodes a protein highly homologous to yeast nuclear protein NHP2. Cytogenet Cell Genet. 72:191-3.

Sleeman, J.E., P. Ajuh, and A.I. Lamond. 2001. snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. J Cell Sci. 114:4407-19.

Sleeman, J.E., and A.I. Lamond. 1999. Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. Curr Biol. 9:1065-74.

Snaar, S., K. Wiesmeijer, A.G. Jochemsen, H.J. Tanke, and R.W. Dirks. 2000. Mutational analysis of fibrillarin and its mobility in living human cells. J Cell Biol. 151:653-62.

Swedlow, J.R., and A.I. Lamond. 2001. Nuclear dynamics: where genes are and how they got there. Genome Biol. 2:REVIEWS0002.

Verheggen, C., J. Mouaikel, M. Thiry, J.M. Blanchard, D. Tollervey, R. Bordonne, D.L. Lafontaine, and E. Bertrand. 2001. Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. Embo J. 20:5480-90.

Vidovic, I., S. Nottrott, K. Hartmuth, R. Luhrmann, and R. Ficner. 2000. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol Cell. 6:1331-42.

Watkins, N.J., A. Dickmanns, and R. Luhrmann. 2002. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. Mol Cell Biol. 22:8342-52.

Watkins, N.J., V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, M. Wilm, M. Rosbash, C. Branlant, and R. Luhrmann. 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. Cell. 103:457-66.

Weinstein, L.B., and J.A. Steitz. 1999. Guided tours: from precursor snoRNA to functional snoRNP. Curr Opin Cell Biol. 11:378-84.

Weis, K. 2002. Nucleocytoplasmic transport: cargo trafficking across the border. Curr Opin Cell Biol. 14:328-35.

CHAPTER VI

PERSPECTIVES

"[The characterisation of the human nucleolar proteome] is a framework for future work

but it doesn't really answer any single question."

~ Tom Misteli, Science 18th January, 2002~

******

"The 1999 concept was the 'dynamic nucleus'.

The 2000-2001 chapter asked, 'the dynamics of what?'.

The 2002 question is how are these dynamics regulated?"

~ Thoru Pederson, Nature cell biology December 2002~

VI Perspectives

VI.1 Summary of findings

In this thesis I have presented 3 aspects of my PhD work on studying the human nucleolus, namely, the characterisation of the human nucleolar proteome (Chapter III), the relationship between subnucleolar domains during mitotic breakdown and reassembly (Chapter IV) and a novel nucleolar targeting pathway (Chapter V). Here is a summary of the major findings:

- 400 proteins were identified that copurified with human nucleoli isolated from HeLa cell nuclei using mass spectrometry and ~30% of these represent proteins that are either novel, or else previously uncharacterised (Andersen et al., 2002);

- Bioinformatics analyses were performed to integrate information from the nucleolar protein with the existing public databases and a dedicated, web-based, multi-platform database was built for public access at http://www.dundee.ac.uk /lifesciences/lamonddatabase/;

- Secondary information, such as mRNA expression levels in 41 human tissues and 115 growth conditions and protein homologues in 80 organisms were derived for over 400 nucleolar proteins and used to classify these proteins, especially the novel factors, for a possible relation to biosynthesis and function of the ribosome *in silico*;

- A battery of 24 HeLa cell lines stably expressing one or more nucleolar factors labelling distinct subdomains were generated and used to study the disassembly and reassembly of the nucleolus in live cells;

- There exist a window between prophase and metaphase when a core subunit of RNA polymerase I does not associate with the chromosomes in live human cells.

This finding is in contrast with previous fixed cell studies suggesting that RNA polymerase I associates with the chromosomes at all stages of mitosis;

- An *in vivo* technique was developed to label one or more nucleoli non-invasively;

- A novel nucleolar localization pathway, observed in primary and transformed cell lines, was discovered for NHPX, a new member that is localised in both the nucleolus and Cajal bodies (Leung and Lamond, 2002);

- The NHPX pathway involves an unexpected transient interaction with splicing speckles prior to nucleolar accumulation;

- The pathway is dependent on pre-mRNA transcription and possibly links several parallel RNA metabolic pathways that occur in distinct subnuclear domains.

VI.2 On Post-Proteomics

The identification of the human nucleolar proteome itself may not, perhaps, answer any single question at the moment but it definitely provides a firm foundation to test new ideas and refute/confirm hypotheses based on these new data (Couzin, 2002; Dundr and Misteli, 2002; Pederson, 2002b). For example, although the idea of mRNA translation in the nucleus is still very controversial (Iborra et al., 2001; Pederson, 2001a), even back in the 1960s, there was data showing that amino acids could be incorporated within isolated nucleoli (Birnstiel and Hyde, 1963; Birnstiel and Flamm, 1964; Maggio, 1966). The data described in this thesis and other studies shows that the presence in the nucleolus of translation factors, ribonucleoparticles/RNAs related to translation, such as SRP, tRNA and 5S RNA, and nearly all the ribosomal proteins, which in total accounting for ~16% of the human nucleolar proteome (Pederson and Politz, 2000; Politz et al., 2000; Andersen et al., 2002; Politz et al., 2002). Of course, the mere presence of these factors in nucleoli does

not necessarily mean that they function as a translation complex, as in the cytoplasm. One way to address the issue of nucleolar translation would be to repeat the amino acid incorporation experiment on isolated human nucleoli, except by using isotopically labeled amino acids. With the aid of the high sensitivity of mass spectrometry, perhaps, it may be possible to identify which proteins the amino acids are incorporated (Flory et al., 2002).

Surprisingly, 30% of the nucleolar proteins are either uncharacterised or novel. This partly reflects the poor status of the current human genome annotation and also presents a challenge for us to unravel the unknown functional complexity of the nucleolus (Andersen et al., 2002). However, if we pool information from different sources such as gene expression activity and information on homologues in other species, as demonstrated in Chapter III, this approach should help speed up our understanding of the human nucleolar proteome (Eisenberg et al., 2000; Marcotte, 2000; Vidal, 2001). The understanding of ribosome biogenesis from budding yeast provides testing materials for the human system (Fatica and Tollervey, 2002), especially now with their corresponding human homologues being mapped in the nucleolus (Chapter III). These putative ribosome biogenesis factors should then be tested rigourously by biochemical methods and these studies are currently underway.

As proteins rarely work alone within the cell, nucleolar proteins may work in a pair or in a complex to perform a specified function. To understand the possible networks within the nucleolus, each nucleolar protein was profiled and compared with one another in this thesis, according to (a) how they are inherited in different genomes, (b) where they are expressed and (c) how they are expressed (c.f. Chapter III). Based on this method, I have classified the possible relationship of each nucleolar protein with the ribosome *in silico*. In particular, 9 novel/uncharacterised proteins were classified in this way and 7 of them have a homologue in yeast that is known to be related to the ribosome.

This profiling technique by incorporating a multitude of information deposited in public databases is a beginning to study biological process modularly, rather than individually (Hartwell et al., 1999; Pellegrini et al., 1999; Eisenberg et al., 2000). The fortuitous discovery of nucleolar functions by studying a single protein requires a more systematic way to analyse its interaction partners, either along a pathway, or within a complex (Tong et al., 2002). Even though the number of the human nucleolar proteins is of the order of hundreds, it is still relatively simple to use computational methods in mapping its potential relationship with one another and studying possible functions within the proteome by mining the wealth of existing and continually expanding databases. However, this is not possible unless efforts are continuously exerted to develop good database structures and user-friendly interfaces (Aebersold and Watts, 2002; Stein, 2002). Good visualisation tools are also important especially for non-specialists from other fields to use the existing data. Moreover, standardization of the protocol for archiving data is crucial for transferring information between laboratories and this procedure can be modeled on existing systems provided, for example, by NCBI and EBI for genomic data and OME for microscopic data (http://ncbi.nlm.nih.gov/; http://www.ebi.ac.uk/; http://www.openmicroscopy.org/).

The nucleolar proteome is nothing but dynamic and therefore the 400 human nucleolar proteins identified forms the basis to understand how the nucleolus responds to change in metabolic environments. As demonstrated by our group, the inhibition of transcription, i.e. by Actinomycin D, causes an enrichment of a set of 11 proteins within the nucleolus (Fox et al., 2002). The advance in mass spectrometry allows quantitative measurement of the relative protein abundance (Flory et al., 2002). This extended capability will certainly add another dimension to our understanding of how the composition of the nucleolus changes under different metabolic states and at specific cell cycle stages. This approach should also

be extended to other cell types, such as stem cells, or those cells, e.g. from brain, where alternative rRNA processing is known to occur (Akhmanova et al., 2000). Mass spectrometry has the added advantage of possibly identifying alternatively spliced forms of the nucleolar proteins. Moreover, quantitative, and even simply qualitative, comparisons between normal and tumour tissues might give insights into the mechanism behind the extremely large nucleoli observed in malignant cells (Derenzini et al., 2000).

Another way of looking at changes to the nucleolar proteome would be to carry out parallel studies using microarray techniques. Gene chips can be tailor-made to contain the 400 corresponding nucleolar cDNAs and study their changes in transcriptional level under different growth conditions. Correlative studies with the quantitative measurement of the relative protein abundance indicated from proteomic data may provide insights into possible post-transcriptional regulation of the nucleolar gene. Moreover, the panel of HeLa cell lines generated during this study that label multiple subcompartments within the nucleolus provides a good model system to study changes in transcription and protein levels as well as at the cellular level under different growth conditions.

VI.3 On Nucleolar structural integrity

As discussed in Chapter IV, the human nucleolus disassembles in the beginning of mitosis at a similar time when the nuclear envelope breaks down. Although nucleolar breakdown was studied between the 1970s and the early 1990s, most of the studies were performed in fixed cells (Lafontaine, 1968; Gautier et al., 1992; Shaw and Jordan, 1995), and therefore these experiments could not possibly identify the inter-relationships between each subnucleolar compartment within a single cell in a temporal sequence. In contrast, this thesis study may give, perhaps, the first insight regarding how the 3 subnucleolar domains move in relation with one another, as well as with the mitotic chromosomes in live cells. It is apparent that chromosome condensation does not play a major role in disrupting the

nucleolus but the breakdown is due to a rapid event that happens concurrently with the nuclear envelope breakdown. So, what is the molecular event that triggers nucleolar breakdown? Because nucleoli can be isolated biochemically as structurally intact and functional entities (Cheutin et al., 2002), this property makes the isolated nucleolus a good *in vitro* system to test which components, or complex, in the mitotic extract triggers the breakdown.

Nucleoli are assembled on the NORs located in the five human chromosome pairs; however, it is often observed that there are less than 5 nucleoli within human cells, so how do these chromosomes arrange themselves among different nucleoli? This study illustrates that the nucleolar reformation initiates stochastically by relocation of RNA polymerase I and subsequent recruitment of other processing and assembly factors as newly pre-rRNA transcripts are made. The *in vivo* labelling technique developed as described in Chapter IV can also be used to label pre-mitotic chromosomal regions surrounding a single nucleolus and follow them as the cells progress through mitosis (e.g. Figure VI-1). Preliminary experiment suggests that the labelled chromosomal regions appear to be segregated to different nucleoli after mitosis and they do not reform to a single nucleolus even after 3 hours of mitosis (data not shown). Therefore future analysis may focus on how these

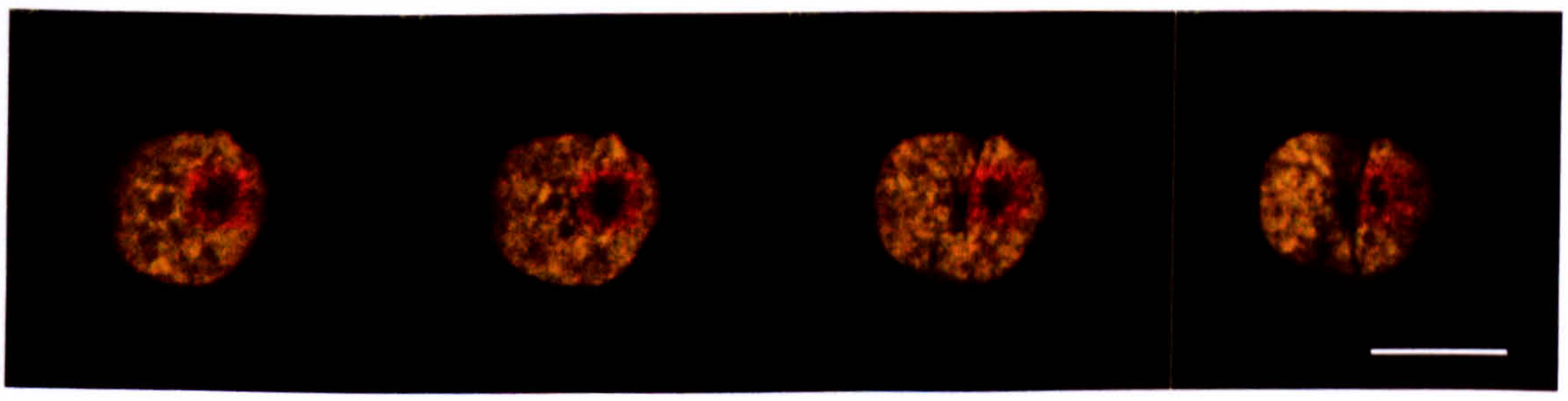


Figure VI-1 ***In vivo* labelling of a nucleolus during mitosis.** HeLa cells stably expressing EYFP-RPA39 and ECFP-H2B were transiently transfected with diHcRed-H2B. A chromosomal region surrounding a single nucleolus was photobleached using a 413 laser for ECFP with a square mark. Shown here is a prophase cell undergoing chromosomal condensation. The images shown here from left to right are a series of optical sections of 1.2μm in the z-axis and only the signals for H2B are shown to illustrate the bleached mark. Yellow: overlapped signals from ECFP-H2B and diHcRed-H2B and Red: bleached region with signals from diHcRed-H2B only. Scale bar = 10μm.

NOR-containing chromosomes are rearranged during the cell cycle, possibly using the lac operon-lacI-GFP system (Belmont and Straight, 1998). This is especially interesting as recent studies suggested that even a transcriptionally silent NOR-containing chromosome can be recruited to the nucleolus (Sullivan et al., 2001).

VI.4 On Nucleolar dynamics and non-dynamics

Another interesting aspect of nucleoli is how this organelle is maintained, to an extent that it can be isolated as a functional complex, if most of the components studied so far are continually exchanged with the nucleoplasm (Shopland and Lawrence, 2000; Misteli, 2001a; Misteli, 2001b; Pederson, 2001b). Indeed, the only way to explain the apparent 'stability' of this and other subnuclear structures is to 'envision that their composition at any given time includes so many stabilised molecules that the shuttling fraction does not materially affect the very existence of the particle' (Pederson, 2002a). For example, two components of PML bodies, PML and sp100 do not shuttle between the nucleoplasm and the PML body, while a third component, CBP, does (Boisvert et al., 2001). Could there be any similar non-shuttling factors to be found in the nucleolus?

It is also possible that RNA molecules provide a structural framework for subnuclear structures. *Xist* RNA has recently been shown to play a structural role in stablising the inactive X chromosome (Clemson et al., 1996). The abundance of non-coding poly(A)+ RNA retained in splicing speckles within the nucleus has been speculated to have a structural role (Huang et al., 1994). Similarly, Cajal bodies have been shown to associate with U2 gene loci, surprisingly via its own transcript, U2 snRNA (Frey and Matera, 2001). Although a lot of nucleolar RNA species are currently known, what is needed is a further systematic analysis of RNAs from isolated human nucleoli. Any such study should especially focus on smaller RNAs, considering the recent expansion of a new RNA class

known as microRNAs which are of 21-23 nucleotides in length (Ambros, 2001; McManus and Sharp, 2002; Moss, 2002).

Or is it the presence of a multitude of different protein-protein or protein-RNA interaction modules that holds the nucleolus together? It should be noted that about 10% of the human nucleolar proteome contains the RNA recognition motif RRM and, more surprisingly, 30% of the nucleolar proteins contains at least one RR[GR] motif (Chapter III). Potentially, the arginine residue in the RG sequences can be methylated and many RNA binding proteins such as SmB, SmD1, SmD3, Lsm5, fibrillarin, hnRNP U and hnRNP Q contain this motif (Paushkin et al., 2002). It is possible that this motif may be mediating interaction with other protein and/or RNA molecules. Moreover, are the 5% of nucleolar proteins containing the protein-protein interaction WD domain involved in binding other nucleolar proteins? Perhaps, nucleolar targeting is more like what happens in the ER where, instead of a particular signal peptide targeting proteins to the ER, proteins that flows away to the Golgi apparatus are retrieved back to the ER by the retention signal KDEL through a shuttling receptor. For example, viral proteins such as HIV-1 Tat and Rev are known to piggy-back to the nucleolus through the binding of the molecular chaperone B23 that shuttles between the nucleoplasm and the nucleolus (Hiscox, 2002).

I proposed in Chapter IV that certain 'adhesive elements', which could be either RNAs or proteins, hold the nucleolus together. If the appearance of the nucleolus is a direct effect of ongoing rRNA transcription, it is difficult to explain why the different subnucleolar domains still contain their usual components when transcription is inhibited, rather than dispersing to the nucleoplasm. So, what holds these factors in distinct subnucleolar domains, even without apparent transcriptional activity going on? The pre-rRNA transcript may be the initial nucleation factor for nucleolar assembly, but there may exist another mechanism to stabilise the nucleolus once it is formed. As discussed in Chapter IV, the

formation of the nucleolus may be similar to a mechanism where cells sort themselves into distinctive clusters according to the differential adhesiveness of different cell types (Steinberg, 1964; Armstrong, 1989; Steinberg and Takeichi, 1994). Different subnucleolar domains may have their own set of 'adhesive elements' and hence arrange themselves into the most stable, and hence functionally efficient, configuration. Understanding the mechanism behind this may also help explain the change in the number and size of nucleoli, and other merging and splitting events observed in other subnuclear structures.

'Dynamics' of nuclear proteins has come to be synonymous with photobleaching experiments using FP-fusion protein expressed in live cell. Re-examination of the recent FRAP data collected for GFP fusions to nucleolar proteins revealed several interesting facts (Table VI-1). First, the transcription inhibition of RNA polymerase I by a low dose of Actinomycin D causes an increase in the diffusion coefficients of fibrillarin, nucleolin, Rpp29, B23 and RPS5, but not the transcription factor UBF (Dundr et al., 2000; Phair and Misteli, 2000; Snaar et al., 2000; Chen and Huang, 2001; Dundr et al., 2002). Second, there exist ~10-20% immobile fractions of the nucleolar factors in the interphase nucleolus, but very tiny, if any, immobile fractions in the nucleoplasm. Third, the immobile fractions of some RNA polymerase I subunits measured are 0% during mitosis compared with 20% in the interphase (Dundr et al., 2002). This is particularly revealing because in this study I observed that the RNA polymerase I is dissociated from the mitotic chromosomes between prophase and metaphase.

Fourth, the data obtained for some proteins are different from one another in two independent studies; for example, both the diffusion coefficient *D* and notably the immobile fraction measured for GFP-fibrillarin differ by two-fold (Phair and Misteli, 2000; Snaar et al., 2000). The most notable difference, apart from using different cell lines, is that

Protein	Conditions		Cell line	Nucleolus			Nucleoplasm			Ref
				$t_{1/2}$	D	Immobile	$t_{1/2}$	D	Immobile	
GFP-UBF1	I	Crtl	HeLa	9	0.13667	0	2.8	0.156786	0	[1]
GFP-UBF1	I	ActD	HeLa	25	0.05684	0.05	2.8	0.51786	0	[1]
GFP-UBF1	I	Crtl	CMT3**	5	ND	0	ND	ND	ND	[2]
GFP-UBF1	I	Crtl	CMT3**	5	ND	0	ND	ND	ND	[2]
GFP-UBF1	M	Crtl	CMT3**	5	ND	0	ND	ND	ND	[2]
GFP-UBF2	I	Crtl	CMT3**	5	ND	0	ND	ND	ND	[2]
GFP-TAF48	I	Crtl	CMT3**	3	ND	0.1	ND	ND	ND	[2]
GFP-PAF53	I	Crtl	CMT3**	10	ND	0.2	ND	ND	ND	[2]
GFP-PAF53	M	Crtl	CMT3**	<5	ND	0.1	ND	ND	ND	[2]
GFP-RPA194	I	Crtl	CMT3**	10	ND	0.1	ND	ND	ND	[2]
GFP-RPA194	I	Crtl	CMT3**	10	ND	0.2	ND	ND	ND	[2]
GFP-RPA194	M	Crtl	CMT3**	<5	ND	0.05	ND	ND	ND	[2]
RPA43-GFP	I	Crtl	CMT3**	10	ND	0.15	ND	ND	ND	[2]
RPA43-GFP	I	Crtl	CMT3**	10	ND	0.2	ND	ND	ND	[2]
RPA43-GFP	M	Crtl	CMT3**	5	ND	0.1	ND	ND	ND	[2]
RPA40-GFP	I	Crtl	CMT3**	10	ND	0.15	ND	ND	ND	[2]
RPA16-GFP	I	Crtl	CMT3**	10	ND	0.2	ND	ND	ND	[2]
FIB-GFP	I	Crtl	BHK	20	0.046	0.25	3	0.53	0	[5]
FIB-GFP	I	Crtl	U2OS*	10	0.02	0.47	ND	ND	ND	[4]
FIB-GFP	I	23°C	BHK	ND	ND	ND	3	0.53	0	[5]
FIB-GFP	I	- ATP	BHK	ND	ND	ND	3	0.53	0	[5]
FIB-GFP	I	ActD	BHK	10	ND	0.2	ND	ND	ND	[5]
FIB-GFP	I	PNB	CMT3	3	ND	0.1	ND	ND	ND	[3]
FIB-GFP	I	NDF	CMT3	3	ND	0.1	ND	ND	ND	[3]
GFP-nucleolin	I	Crtl	HeLa	9	0.14444	0	1.3	1.15385	0	[1]
GFP-nucleolin	I	ActD	HeLa	8	0.152941	0	1	1.34657	0	[1]
GFP-Rpp29	I	Crtl	HeLa	18	0.075	0.05	1	1.5436	0	[1]
GFP-Rpp29	I	ActD	HeLa	8	0.164706	0	1.4	1.25833	0	[1]
GFP-B23	I	Crtl	HeLa	20	0.0785	0	1	1.51	0	[1]
GFP-B23	I	ActD	HeLa	17	0.085294	0	1	1.56748	0	[1]
GFP-S5	I	Crtl	HeLa	72	0.019028	0.4	1.2	1.258333	0	[1]
GFP-S5	I	ActD	HeLa	32	0.042813	0.2	1	1.59	0	[1]

Table VI-1 **Summary of FRAP data of GFP-nucleolar components examined so far.** 'Conditions' indicates the corresponding cell cycle state and metabolic state: I for interphase and M for metaphase; Crtl: control, 23°C: cells grown at 23°C, -ATP: cells depleted of ATP and ActD: 0.04-0.05μg/ml of the transcription inhibitor. Actinomycin D was used at this concentration in these studies to specifically inhibit RNA polymerase I. 'Cell line' indicates the corresponding cell line: all results are obtained from transient transfections unless specified otherwise; * indicates stable transfection and ** indicates the CMT3/HSC/NIH3T3 cell lines were used either stably or transiently transfected (which is not specified in the original publication); $t_{1/2}$ indicates the time required to recover half of the fluorescence loss at the bleached area and some of the results are estimated from the given plots; D: Diffusion coefficient; 'Immobile' indicates the immobile fraction; ND = not determined. See Chapter I for the principle of FRAP. References (Ref): [1] Chen and Huang, 2001; [2] Dundr et al, 2002; [3] Dundr et al, 2000; [4] Snaar et al, 2000 and [5] Phair and Misteli, 2000.

the case that shows a smaller *D* and a larger immobile fraction (47%) used a stably transfected GFP-fibrillarin cell line. The discrepancy may thus be explained by the presence of higher protein level, due to the transient expression of exogenous protein in one case,

which could raise the apparent mobility if endogenous binding sites become saturated. Moreover, my FRAP data on RNA polymerase I subunit RPA39, using a stably transfected HeLa cell line, shows a much longer time for the cell to recover half of the fluorescence loss after photobleaching (~100s) compared with 3s obtained by another group, in which the report did not specify whether the experiments were performed on transiently or stably transfected cells (Dundr et al., 2002). Therefore, it would be better to consider the effect of 'overexpression' with regard to the number of potential binding sites when interpreting the current results (Shopland and Lawrence, 2000), and, more importantly, confirm these results by other independent methods.

Another problem of FRAP and other photobleaching techniques is that they cannot tell the molecular status of the protein represented by the data. The result is an integral of free or bound, activated or repressed and monomeric or complexed forms of the molecule of interest (Chen and Huang, 2001; Lippincott-Schwartz et al., 2001). Therefore, it may not truly represent the diffusion rate of the substrate-bound active form of the molecule within a subnuclear structure, e.g. the nucleolus. Moreover, to what extent can we interpret the current existing data which contain an immobile fraction within the nucleolus, but not in the nucleoplasm, as dynamic? The fact that the protein can be dynamically exchanged between the nucleolus and the nucleoplasm does not exclude the possibility that a fraction of the molecule (47% in one case for fibrillarin) is either bound to an intranucleolar, fixed/stable structure, or else forms large complexes that are restricted in movement.

VI.5 On Integration of expression pathways

Although subnuclear structures are not separated by membranes, it has been demonstrated in this thesis and other studies that directional pathways between structures exist for nuclear proteins (Sleeman and Lamond, 1999a; Leung and Lamond, 2002). It is important to understand what causes these nuclear proteins to enter a structure? or to leave a

structure? One way to understand these is to identify the molecular interaction partners of these proteins when they are localised in distinct domains.

NHPX is known to be the initial nucleation factor that binds the RNA for both snoRNP and snRNP (Nottrott et al., 2002; Watkins et al., 2002). One of the possibilities is that NHPX changes its RNA partner as it progresses along the pathway. Spliceosomal snRNA is known to be exported to the cytoplasm for hypermethylation and is re-imported into the nucleus whereas the hypermethylation of snoRNAs occurs in the nucleus, not the cytoplasm (Nigg et al., 1991; Sleeman and Lamond, 1999b). Therefore, the hypermethylated form of U4 has a physical advantage to interact with the newly expressed NHPX in the cytoplasm before their nuclear entry and hence subsequent localisation in the splicing speckles.

But what could possibly trigger the change of RNA partners? I have shown that the progression of the NHPX pathway requires ongoing pre-mRNA transcription. As discussed in chapter V, the progression may rely on splicing of those pre-mRNAs whose introns contain pre-snoRNAs (Weinstein and Steitz, 1999). It is feasible that the U4 snRNP is disassembled after each round of splicing, and the factors, which form the complex, e.g. NHPX, are recycled. Therefore, NHPX may be released from the once-packed complex and is now able to bind to the nascent snoRNA (Hirose and Steitz, 2001). As all snoRNAs in vertebrates are housed within the introns of a family of 5' TOP pre-mRNA transcripts (see Chapter I), whose products are commonly ribosomal proteins or factors related to ribosomal biogenesis (Weinstein and Steitz, 1999), this model may link the splicing of gene transcripts that are related to ribosomal biogenesis to the production of snoRNAs that guide the modification of rRNAs. To test this model, *in vitro* assembled NHPX-U3 and -U4 RNPs can be microinjected into HeLa cells to examine and contrast their localisation and hence test whether the localisation is dependent on the RNA species

NHPX binds to. Moreover, it would be interesting to test whether splicing or transcription is more important for the progression of the NHPX pathway by inhibiting splicing rather than the transcription of pre-mRNA specifically *in vivo*.

Another interesting aspect of this pathway is the continual presence of NHPX in Cajal bodies, both at the early and late stage, although the pathway still progresses in cell lines lacking prominent Cajal bodies. As gems and Cajal bodies are often colocalised within human nuclei, it would be interesting to see which subnuclear compartment NHPX would go to at different stages in a cell line such as HeLa PV, that shows a separation of Cajal bodies and gems, or in coilin-knockout mouse cells (Liu and Dreyfuss, 1996; Tucker et al., 2001). These approaches should give insights into understanding which protein components in Cajal bodies/gems are interacting with NHPX during the pathway progression.

VI.6 Final words

As described by Valentin back in 1836, the nucleolus is like 'a secondary nucleus within the nucleus'. In fact, its complexity revealed by its subcompartmentalisation may provide a good model to study how the nucleus is organised. The number of nucleolar proteins (~400-500) is ideal for identifying modular sets of proteins which perform specific functions by large-scale profiling of the proteins under many different conditions. The opportunity to isolate functionally intact nucleoli using a well-established protocol, possibly from the established HeLa cell lines with fluorescently labeled subcompartments, provides a reproducible *in vitro* system to test models both biochemically and microscopically. The current understanding of the human nucleolar proteome and dynamics may be nothing but a teaser trailer, yet we are all welcome to the show.

VI.7 References

Aebersold, R., and J.D. Watts. 2002. The need for national centers for proteomics. *Nat Biotechnol.* 20:651.

Akhmanova, A., T. Verkerk, A. Langeveld, F. Grosveld, and N. Galjart. 2000. Characterisation of transcriptionally active and inactive chromatin domains in neurons. *J Cell Sci.* 113 Pt 24:4463-74.

Ambros, V. 2001. microRNAs: tiny regulators with great potential. *Cell.* 107:823-6.

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. *Curr Biol.* 12:1-11.

Armstrong, P.B. 1989. Cell sorting out: the self-assembly of tissues in vitro. *Crit Rev Biochem Mol Biol.* 24:119-149.

Belmont, A.S., and A.F. Straight. 1998. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends Cell Biol.* 8:121-4.

Birnstiel, M., and W. Flamm. 1964. Intranuclear site of histone synthesis. *Science.* 145:1435-1437.

Birnstiel, M., and B. Hyde. 1963. Protein synthesis by isolated pea nucleoli. *J Cell Biol.* 18:41-50.

Boisvert, F.M., M.J. Kruhlak, A.K. Box, M.J. Hendzel, and D.P. Bazett-Jones. 2001. The transcription coactivator CBP is a dynamic component of the promyelocytic leukemia nuclear body. *J Cell Biol.* 152:1099-106.

Chen, D., and S. Huang. 2001. Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. *J Cell Biol.* 153:169-76.

Cheutin, T., M.F. O'Donohue, A. Beorchia, M. Vandelaer, H. Kaplan, B. Defever, D. Ploton, and M. Thiry. 2002. Three-dimensional organization of active rRNA genes within the nucleolus. *J Cell Sci.* 115:3297-307.

Clemson, C.M., J.A. McNeil, H.F. Willard, and J.B. Lawrence. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol.* 132:259-75.

Couzin, J. 2002. Stripping the Nucleolus Down to Its Proteins. *Science.* 295:422.

Derenzini, M., D. Trere, A. Pession, M. Govoni, V. Sirri, and P. Chieco. 2000. Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. *J Pathol.* 191:181-6.

Dundr, M., U. Hoffmann-Rohrer, Q. Hu, I. Grummt, L.I. Rothblum, R.D. Phair, and T. Misteli. 2002. A kinetic framework for a mammalian RNA polymerase in vivo. *Science.* 298:1623-6.

Dundr, M., and T. Misteli. 2002. Nucleolomics: an inventory of the nucleolus. *Mol Cell.* 9:5-7.

Dundr, M., T. Misteli, and M.O. Olson. 2000. The dynamics of postmitotic reassembly of the nucleolus. *J Cell Biol.* 150:433-46.

Eisenberg, D., E.M. Marcotte, I. Xenarios, and T.O. Yeates. 2000. Protein function in the post-genomic era. *Nature.* 405:823-6.

Fatica, A., and D. Tollervey. 2002. Making ribosomes. *Curr Opin Cell Biol.* 14:313-8.

Flory, M., T. Griffin, D. Martin, and R. Aebersold. 2002. Advances in quantitative proteomics using stable isotope tag. *Trends in Biotechnology.* 20:S23-S29.

Fox, A.H., Y.W. Lam, A.K. Leung, C.E. Lyon, J. Andersen, M. Mann, and A.I. Lamond. 2002. Paraspeckles. A novel nuclear domain. *Curr Biol.* 12:13-25.

Frey, M.R., and A.G. Matera. 2001. RNA-mediated interaction of Cajal bodies and U2 snRNA genes. *J Cell Biol.* 154:499-509.

Gautier, T., M. Robert-Nicoud, M.N. Guilly, and D. Hernandez-Verdun. 1992. Relocation of nucleolar proteins around chromosomes at mitosis. A study by confocal laser scanning microscopy. *J Cell Sci.* 102 (Pt 4):729-37.

Hartwell, L.H., J.J. Hopfield, S. Leibler, and A.W. Murray. 1999. From molecular to modular cell biology. *Nature.* 402:C47-52.

Hirose, T., and J.A. Steitz. 2001. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc Natl Acad Sci U S A.* 98:12914-9.

Hiscox, J.A. 2002. The nucleolus--a gateway to viral infection? *Arch Virol.* 147:1077-89.

Huang, S., T.J. Deerinck, M.H. Ellisman, and D.L. Spector. 1994. In vivo analysis of the stability and transport of nuclear poly(A)+ RNA. *J Cell Biol.* 126:877-99.

Iborra, F.J., D.A. Jackson, and P.R. Cook. 2001. Coupled transcription and translation within nuclei of mammalian cells. *Science.* 293:1139-42.

Lafontaine, J. 1968. Structural components of the nucleus in mitotic plant cells. *In* The Nucleus. A. Dalton and F. Hagenau, editors. Academic Press, New York. 152-96.

Leung, A.K., and A.I. Lamond. 2002. In vivo analysis of NHPX reveals a novel nucleolar localization pathway involving a transient accumulation in splicing speckles. *J Cell Biol.* 157:615-29.

Lippincott-Schwartz, J., E. Snapp, and A. Kenworthy. 2001. Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol.* 2:444-56.

Liu, Q., and G. Dreyfuss. 1996. A novel nuclear structure containing the survival of motor neurons protein. *Embo J.* 15:3555-65.

Maggio, R. 1966. Progress report on the characterization of nucleoli from guinea pig liver. *Nat. Cancer Inst. Monograph.* 23:213-222.

Marcotte, E.M. 2000. Computational genetics: finding protein function by nonhomology methods. *Curr Opin Struct Biol.* 10:359-65.

McManus, M.T., and P.A. Sharp. 2002. Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet.* 3:737-47.

Misteli, T. 2001a. The concept of self-organization in cellular architecture. *J Cell Biol.* 155:181-5.

Misteli, T. 2001b. Protein dynamics: implications for nuclear architecture and gene expression. *Science.* 291:843-7.

Moss, E.G. 2002. MicroRNAs: hidden in the genome. *Curr Biol.* 12:R138-40.

Nigg, E.A., P.A. Baeuerle, and R. Luhrmann. 1991. Nuclear import-export: in search of signals and mechanisms. *Cell.* 66:15-22.

Nottrott, S., H. Urlaub, and R. Luhrmann. 2002. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *Embo J.* 21:5527-38.

Paushkin, S., A.K. Gubitz, S. Massenet, and G. Dreyfuss. 2002. The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol.* 14:305-12.

Pederson, T. 2001a. Is the nucleus in need of translation? *Trends Cell Biol.* 11:395-7.

Pederson, T. 2001b. Protein mobility within the nucleus--what are the right moves? *Cell.* 104:635-8.

Pederson, T. 2002a. Dynamics and genome-centricity of interchromatin domains in the nucleus. *Nat Cell Biol.* 4:E287-91.

Pederson, T. 2002b. Proteomics of the nucleolus: more proteins, more functions? *Trends Biochem Sci.* 27:111-2.

Pederson, T., and J.C. Politz. 2000. The nucleolus and the four ribonucleoproteins of translation. *J Cell Biol.* 148:1091-5.

Pellegrini, M., E.M. Marcotte, M.J. Thompson, D. Eisenberg, and T.O. Yeates. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A.* 96:4285-8.

Phair, R.D., and T. Misteli. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature.* 404:604-9.

Politz, J.C., L.B. Lewandowski, and T. Pederson. 2002. Signal recognition particle RNA localization within the nucleolus differs from the classical sites of ribosome synthesis. *J Cell Biol.* 159:411-8.

Politz, J.C., S. Yarovoi, S.M. Kilroy, K. Gowda, C. Zwieb, and T. Pederson. 2000. Signal recognition particle components in the nucleolus. *Proc Natl Acad Sci U S A.* 97:55-60.

Shaw, P.J., and E.G. Jordan. 1995. The nucleolus. *Annu Rev Cell Dev Biol.* 11:93-121.

Shopland, L.S., and J.B. Lawrence. 2000. Seeking common ground in nuclear complexity. *J Cell Biol.* 150:F1-4.

Sleeman, J.E., and A.I. Lamond. 1999a. Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol.* 9:1065-74.

Sleeman, J.E., and A.I. Lamond. 1999b. Nuclear organization of pre-mRNA splicing factors. *Curr Opin Cell Biol.* 11:372-7.

Snaar, S., K. Wiesmeijer, A.G. Jochemsen, H.J. Tanke, and R.W. Dirks. 2000. Mutational analysis of fibrillarin and its mobility in living human cells. *J Cell Biol.* 151:653-62.

Stein, L. 2002. Creating a bioinformatics nation. *Nature.* 417:119-20.

Steinberg, M.S. 1964. The problem of adhesive selectivity in cellular interactions. *In* Cellular Membranes in Development. M. Locke, editor. Academic Press, New York. 321-366.

Steinberg, M.S., and M. Takeichi. 1994. Experimental specification of cell sorting, tissue spreading, and specific spatial patterning by quantitative differences in cadherin expression. *Proc Natl Acad Sci U S A.* 91:206-9.

Sullivan, G.J., J.M. Bridger, A.P. Cuthbert, R.F. Newbold, W.A. Bickmore, and B. McStay. 2001. Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *Embo J.* 20:2867-74.

Tong, A.H., B. Drees, G. Nardelli, G.D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C.W. Hogue, S. Fields, C. Boone, and G. Cesareni. 2002. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science.* 295:321-4.

Tucker, K.E., M.T. Berciano, E.Y. Jacobs, D.F. LePage, K.B. Shpargel, J.J. Rossire, E.K. Chan, M. Lafarga, R.A. Conlon, and A.G. Matera. 2001. Residual Cajal bodies in coilin knockout mice fail to recruit Sm snRNPs and SMN, the spinal muscular atrophy gene product. *J Cell Biol.* 154:293-307.

Vidal, M. 2001. A biological atlas of functional maps. *Cell.* 104:333-9.

Watkins, N.J., A. Dickmanns, and R. Luhrmann. 2002. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol.* 22:8342-52.

Weinstein, L.B., and J.A. Steitz. 1999. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol.* 11:378-84.

Chapter VI Perspectives

APPENDIX I:

PUBLICATION REPRINTS

In vivo analysis of NHPX reveals a novel nucleolar localization pathway involving a transient accumulation in splicing speckles. *J Cell Biol.* (2002) 157:615-29
Leung AK and Lamond AI.

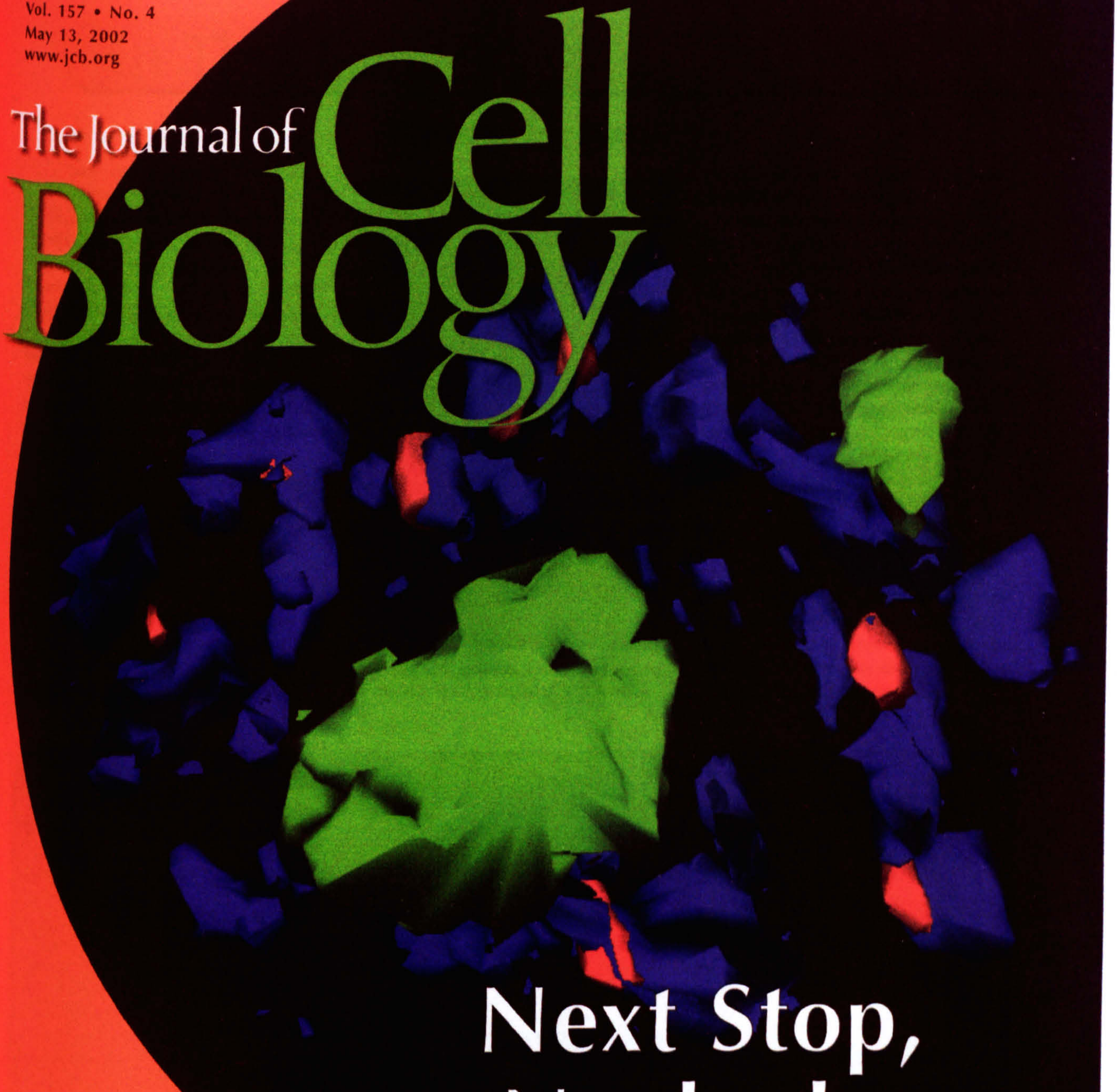
Yeast Pescadillo is required for multiple activities during 60S ribosomal subunit synthesis. *RNA.* (2002) 8:626-36
Oeffinger M, Leung A, Lamond A, Tollervey D.

Directed Proteomic Analysis of the Human Nucleolus. *Curr. Biol.* (2002) 12:1-12
Andersen J, Lyon CE, Fox AH, Leung AK, Lam YW, Steen H, Mann M, Lamond AI.

Paraspeckles: A novel nuclear domain. *Curr. Biol.* (2002) 12:13-25
Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI.

Vol. 157 • No. 4
May 13, 2002
www.jcb.org

The Journal of Cell Biology



Next Stop, Nucleolus

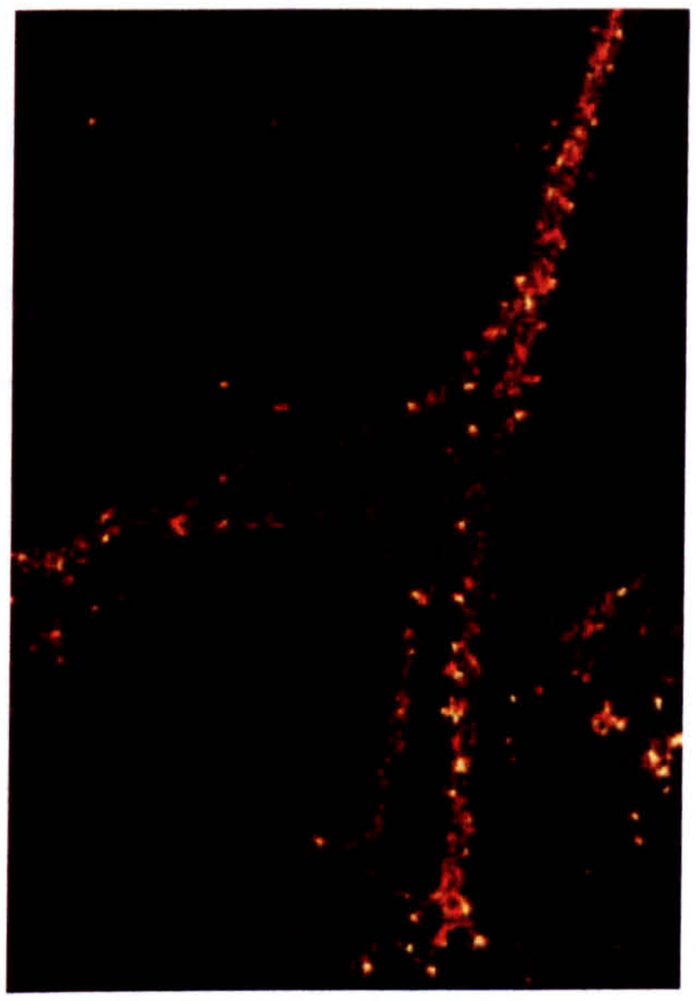
A Stabilizer for Meiosis II Spindles • The MAG Receptor
Spectrin's Muscular Function • Transcription in Tight Spaces

In This Issue

Turn-ons for axons

When adult neurons are injured, myelin-associated glycoprotein (MAG) triggers signals to block the regrowth of axons, a process that is intensely frustrating to patients and researchers hoping to treat such injuries. On page 565, Yamashita et al. identify a receptor complex responsible for transducing this signal, elucidate part of the associated intracellular signaling pathway, and uncover a novel molecular interaction that may help explain the diverse activities of a neurotrophin receptor.

MAG signaling is known to inhibit axonal regrowth after injury, but the MAG receptor on neurons remained unknown. By analyzing neurons from wild-type and mutant mice, the authors discovered that $p75^{NTR}$, a glycoprotein on the surface of many types of neurons, transduces the MAG signal. Since $p75^{NTR}$ also binds to neurotrophins, which promote axonal outgrowth, the same receptor seems to be capable of transducing both growth and inhibitory signals.



Growth-blocking MAG (green) binds to the neurotrophin receptor $p75^{NTR}$ (red).

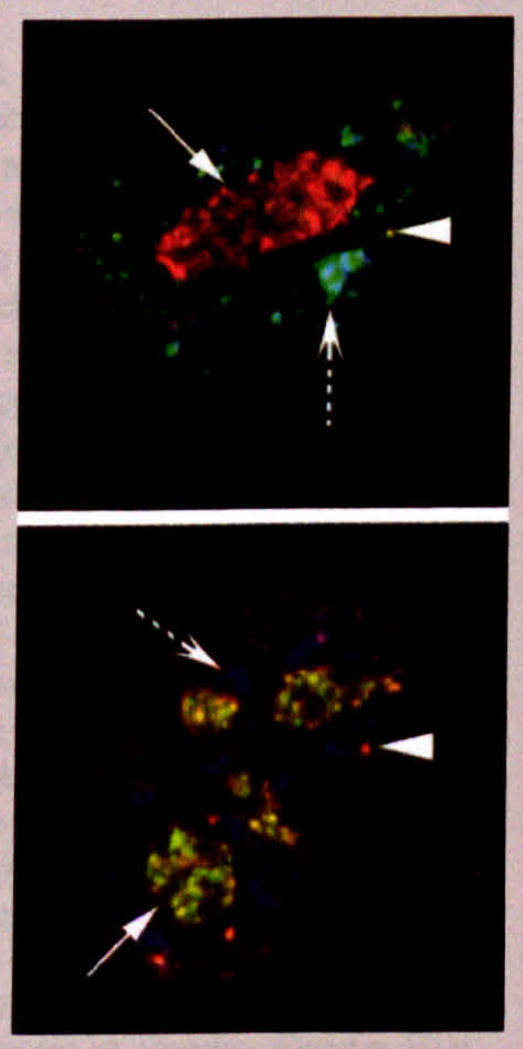
Rather than binding directly to $p75^{NTR}$, MAG interacts with a complex of $p75^{NTR}$ and the ganglioside GT1b. This is the first time a ganglioside has been shown to act as a coreceptor, a finding that may set a precedent for future analysis of these molecules. MAG binding to the complex activates RhoA, which has previously been shown to inhibit neurite extension. The results suggest that GT1b acts primarily as a binding partner, whereas $p75^{NTR}$ is a signal transducing element for MAG.

Since gangliosides are found in lipid rafts, an intriguing possibility is that $p75^{NTR}$ interacts with different factors in a raft to transduce opposite types of signals, depending on its interaction with MAG or neurotrophins. Molecules that target specific components of these pathways might be able to induce desired growth patterns in injured neurons. ■

Nuclear proteins get around

Protein trafficking between the cytoplasm and the nucleus has been studied extensively; but how do proteins move from one site to another within the nucleus? On page 615, Leung and Lamond focus on the intranuclear trafficking of the RNA-binding protein NHPX, and describe an elegant series of experiments that demonstrates the existence of multiple intranuclear accumulation pathways. The study provides definitive evidence of protein sorting within the nucleus.

Although previous studies have suggested that proteins might follow directed paths between intranuclear bodies, the new work provides a detailed analysis of this phenomenon. NHPX localizes primarily to nucleoli, but is capable of binding to both small nucleolar RNAs (snoRNAs) and the spliceosomal U4 small nuclear RNA (snRNA). The authors followed fluorescent fusion proteins to determine that newly expressed NHPX transiently visits splicing speckles in the nucleus before accumulating stably in the nucleolus, a pattern confirmed in multiple cell lines. The move from speckles to nucleolus is apparently unidirectional, and requires new mRNA transcription, suggesting that other factors must be expressed to allow NHPX to leave the speckles.



A nucleolar protein (green) is first (top) found in splicing speckles (blue) before later moving to nucleolar structures (red).

Leung and Lamond also compare the trafficking of NHPX to that of the SmB protein. In contrast to NHPX, SmB accumulates first in Cajal bodies and in the nucleolus, before finally accumulating in speckles, indicating that the two proteins move through distinct sorting pathways with in the nucleus.

The unidirectional movement of NHPX, combined with its previously observed binding properties, suggests that NHPX might interact with U4 snRNA in speckles, possibly facilitating the maturation of nuclear proteins or RNP complexes before moving to the nucleolus. The authors are now trying to determine whether other nuclear proteins share the separate trafficking routes used by NHPX and SmB, and they hope to identify the molecular mechanisms responsible for these pathways. ■

JCB Article

In vivo analysis of NHPX reveals a novel nucleolar localization pathway involving a transient accumulation in splicing speckles

Anthony K.L. Leung and Angus I. Lamond

Wellcome Trust Biocentre, MSI/WTB Complex, University of Dundee, Dundee DD1 5EH, Scotland, UK

The NHPX protein is a nucleolar factor that binds directly to a conserved RNA target sequence found in nucleolar box C/D snoRNAs and in U4 snRNA. Using enhanced yellow fluorescent protein (EYFP)– and enhanced cyan fluorescent protein–NHPX fusions, we show here that NHPX is specifically accumulated in both nucleoli and Cajal bodies (CBs) in vivo. The fusion proteins display identical localization patterns and RNA binding specificities to the endogenous NHPX. Analysis of a HeLa cell line stably expressing EYFP–NHPX showed that the nucleolar accumulation of NHPX was preceded by its transient accumulation in splicing speckles. Only newly expressed NHPX accumulated in speckles, and the nucleolar pool of NHPX did not interchange with the pool in speckles, consistent with a unidirectional pathway. The transient accumulation of NHPX in speckles prior to nucleoli was observed in multiple cell lines, including primary cells that lack CBs. Inhibitor studies indicated that progression of newly expressed NHPX from speckles to nucleoli was dependent on RNA polymerase II transcription, but not on RNA polymerase I activity. The data show a specific temporal pathway involving the sequential and directed accumulation of NHPX in distinct subnuclear compartments, and define a novel mechanism for nucleolar localization.

Introduction

The cell nucleus is the site at which chromosomes are located, and at which DNA replication and gene expression are coordinated and regulated. Many nuclear factors are organized into specific structures termed nuclear bodies (for review see Lamond and Earnshaw, 1998; Schul et al., 1998; Matera, 1999; Spector, 2001). Most nuclear bodies are dynamic, disassembling on entry to M phase and reassembling after mitosis (Dundr et al., 2000). Unlike cytoplasmic organelles, nuclear bodies are not enclosed by membranes. Factors can move in and out of nuclear bodies (for review see Misteli, 2001), and the bodies themselves can also move within the nucleoplasm (Boudonck et al., 1999; Platani et al., 2000; Snaar et al., 2000; Muratani et al., 2001).

The best-studied subnuclear body is the nucleolus (for reviews see Pederson, 1998; Scheer and Hock, 1999; Carmo-Fonseca et al., 2000; Olson et al., 2000; Visintin and Amon, 2000). Whereas ~10% of protein-coding genes are active at any time, ~80% of the total cellular transcription activity is devoted to expression of the conserved ribosomal DNA (rDNA)* repeats in nucleoli by RNA polymerase I (Blobel and Potter, 1967). Multiple rDNA transcription units are joined together by nontranscribed intergenic spaces to form the so-called nucleolar organizing regions (NORs) located in human chromosomes 13, 14, 15, 21, and 22. Nucleoli are assembled on these chromosomal NOR regions after every round of mitosis. The rDNA-containing region is transcribed to yield a precursor, the 45S pre-rRNA, which is processed in a series of posttranscriptional cleavage and modification reactions to generate mature rRNA species that form the catalytic core of the ribosome (for review see Warner, 2001). Various small nucleolar RNAs (snoRNAs), for example U3 and U14, are needed for these separate rRNA cleavage steps, each snoRNA being required at specific steps. The snoRNAs can be categorized into two main classes, box C/D (i.e., U3, U8, U14) and box H/ACA (i.e., U17, U19, U64), according to conserved sequence elements and the way in which they are assumed to fold into defined secondary structures (for review see Weinstein and Steitz,

Address correspondence to Angus I. Lamond, Wellcome Trust Biocentre, MSI/WTB Complex, University of Dundee, Dundee DD1 5EH, Scotland, UK. Tel.: 44-1382-345473. Fax: 44-1382-345695. E-mail: a.i.lamond@dundee.ac.uk

Key words: nucleolus; Cajal bodies; speckles; localization; nucleus

The Journal of Cell Biology, Volume 157, Number 4, May 13, 2002 615–629
http://www.jcb.org/cgi/doi/10.1083/jcb.200201120

*Abbreviations used in this paper: CB, Cajal body; ECFP, enhanced cyan fluorescent protein; EYFP, enhanced yellow fluorescent protein; FIB, fibrillarin; FLIP, fluorescence loss in photobleaching; GFP, green fluorescent protein; NOR, nucleolar organizing regions; PEG, polyethene glycol; rDNA, ribosomal DNA; snoRNA, small nucleolar RNA.

1999). These snoRNAs assemble together with protein factors to form snoRNPs that have roles in guiding and catalyzing posttranscriptional RNA modifications. For example, box C/D snoRNPs direct 2′-*O*-ribose methylation and box H/ACA snoRNPs direct pseudouridylations of specific rRNA nucleotides (for review see Lafontaine and Tollervey, 1998; Lewis and Tollervey, 2000).

Several nucleolar proteins, including the snoRNP factor fibrillarin, have been found in Cajal bodies (CBs) (for review see Gall, 2000). These subnuclear structures, originally called nucleolar accessory bodies, or coiled bodies, are often associated with the nucleolar periphery, or even located inside the nucleolus (Malatesta et al., 1994; Ochs et al., 1994; Lyon et al., 1997; Sleeman et al., 1998). In addition to snoRNPs, they also contain splicing snRNPs and some transcription factors; however, they do not contain non-snRNP protein splicing factors or nascent pre-mRNA. Their snRNP components are specifically newly assembled particles (Carvalho et al., 1999; Gall et al., 1999; Sleeman and Lamond, 1999a; Sleeman et al., 2001). Fluorescently labeled U3, U8, and U14 snoRNAs, when microinjected into *Xenopus* oocytes, accumulate transiently in CBs prior to nucleoli, suggesting that newly imported snoRNAs flow from the CBs to nucleoli (Narayanan et al., 1999b). Therefore, CBs might be involved in transport and maturation of both snRNPs and snoRNPs.

In higher eukaryotes, most pre-mRNAs must be spliced to generate functional mRNAs. Splicing is a nuclear process carried out by a complex RNA protein machine termed the spliceosome (for review see Staley and Guthrie, 1998). Although splicing often occurs cotranscriptionally, the majority of pre-mRNA splicing factors are not localized at active transcription sites; instead, they are enriched in domains called speckles (for reviews see Sleeman and Lamond, 1999b; Misteli, 2000). Recruitment of splicing factors to sites of transcription is believed to involve a cycle of phosphorylation and dephosphorylation (for review see Misteli, 1999).

The advent of green fluorescent protein (GFP) technology allows the visualization of tagged proteins and their movement in living cells (for review see Swedlow and Lamond, 2001). The use of photobleaching techniques with the GFP fusion proteins also allows analysis of protein dynamics (Phair and Misteli, 2001; Reits and Neefjes, 2001). Here we describe a novel nuclear pathway shown by the nucleolar factor NHPX. The NHPX protein was recently identified as a putative human homologue of yeast NHP2p (Saito et al., 1996), and later, independently, as a 15.5-kD RNA-binding protein (Nottrott et al., 1999). Subsequent analyses indicated that NHPX is the human orthologue of the *Saccharomyces cerevisiae* Snu13p (Chang et al., 1999; Nottrott et al., 1999). The human NHPX protein was also independently identified in a directed proteomic analysis of nucleoli isolated from cultured HeLa cells (Andersen et al., 2002); it shares a common RNP structure that binds to both box C/D snoRNAs and spliceosomal U4 snRNA (Nottrott et al., 1999; Vidovic et al., 2000; Watkins et al., 2000). However, NHPX localizes in nucleoli by immunofluorescence (Chang et al., 1999). Here we analyze the in vivo pathway by which NHPX accumulates in nucleoli, and show that it involves an unexpected, transient interaction with splicing speckles. The movement of NHPX from speckles to nucleoli is dependent on pre-mRNA transcription. These data suggest that NHPX may be involved in, and possibly link, several parallel RNA metabolic pathways that occur in distinct nuclear domains.

Results

NHPX is localized to nucleoli and CBs

NHPX was selected for characterization as part of the analysis of the human nucleolar proteome (Andersen et al., 2002). To address its in vivo localization, the NHPX cDNA was isolated from a HeLa cDNA library and tagged at its NH_2 terminus with the enhanced yellow fluorescent protein (EYFP) gene to generate plasmid pAL107$^{EYFP-NHPX}$ (see Materials and methods). The localization of the EYFP–NHPX fusion protein was analyzed by fluorescence microscopy after transient transfection of pAL107$^{EYFP-NHPX}$ in HeLa cells (Fig. 1), and compared with the nucleolar protein fibrillarin (FIB). Transiently expressed EYFP–NHPX was colocalized with FIB in the dense fibrillar component of nucleoli and also, unexpectedly, in CBs. (Fig. 1 A, arrow indicates nucleolus, arrowhead indicates CB). The CB localization was confirmed by counterstaining with anti-coilin antibodies (Fig. 2 A; unpublished data). The localization of NHPX to CBs and nucleoli was also confirmed by immunofluorescence (Figs. 1 B and S1, available at http://www.jcb.org/cgi/content/full/200201120/DC1). The tagged NHPX showed an identical localization to its endogenous counterpart, suggesting that NHPX is a component of both nuclear structures (Fig. 1 B).

To facilitate in vivo analysis of NHPX, we established a HeLa cell line stably expressing EYFP–NHPX (see Materials and methods). As shown by fluorescence microscopy, the EYFP–NHPX protein localized specifically in nucleoli and CBs in the HeLa$^{EYFP-NHPX}$ stable cell line (Fig. 2). The HeLa$^{EYFP-NHPX}$ cell line grows and divides at the same rate as the parental HeLa cell line without any apparent cell cycle defect, as shown by FACS analysis (Fig. 1 C; unpublished data). The expression level of EYFP–NHPX is comparable to that of the endogenous NHPX protein in the HeLa$^{EYFP-NHPX}$ cell line (Fig. 1 D).

To confirm that the EYFP–NHPX fusion protein expressed in the HeLa$^{EYFP-NHPX}$ cell line behaves biochemically as the endogenous protein, its in vivo RNA-binding specificity was investigated. An extract from HeLa$^{EYFP-NHPX}$ cells was immunoprecipitated using an anti-GFP antibody. RNAs in the immunoprecipitate were separated by Urea-polyacrylamide gel electrophoresis, transferred to nylon membrane, and hybridized with probes for U1, U2, U4, U5, and U6 snRNAs and U3 snoRNA (Fig. 1 E). This showed that U3 snoRNA and U4 snRNA, but not U1, U2, U5, or U6 snRNAs, were preferentially isolated with the anti-GFP antibody (Fig. 1 E, lane 3). Control experiments, i.e., bead control and an HeLa$^{EGFP-H2B}$ cell extract, showed that although the anti-GFP antibody still pulled down the fluorescent protein fusion protein (unpublished data), it did not pull down any of the RNAs tested from these extracts (Fig. 1 E, lane 1 and 2), whereas the same anti-GFP antibody pulled down each of the U1, U2, U4, U5, and U6 snRNAs, but not U3 snoRNA, from a HeLa$^{ECFP-SmB}$ cell extract (Fig. 1 E, lane 4). These data show that the EYFP–NHPX fu-

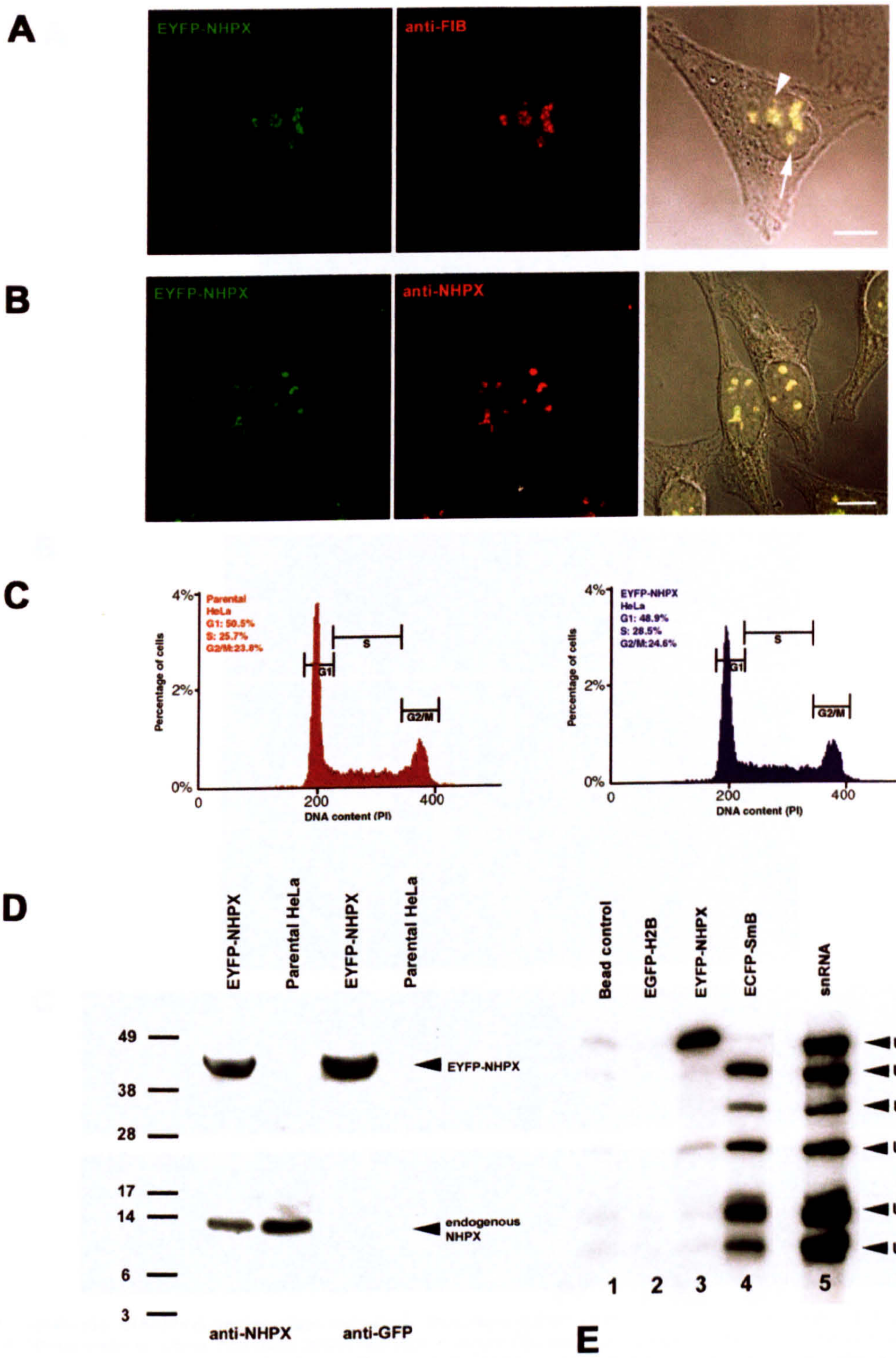


Figure 1. **Transient and stable expression of EYFP–NHPX in HeLa cell line.** HeLa cells transiently transfected with pAL107$^{EYFP-NHPX}$ for 16 h were fixed and counterstained with (A) anti-FIB antibody 72B9 and (B) affinity-purified anti-NHPX serum R86. Arrowhead indicates CB and arrow indicates nucleolus. (C) DNA content of HeLa$^{EYFP-NHPX}$ cells were analyzed by FACS analysis and (D) its expression level by immunoblot using antiserum R86 and anti-GFP. (E) The in vivo RNA binding activity of EYFP-NHPX was assayed by immunoprecipitation and the binding of snRNAs U1, U2, U4, U5, and U6 as well as snoRNA U3 were tested by Northern hybridization. Bars, 5 μm.

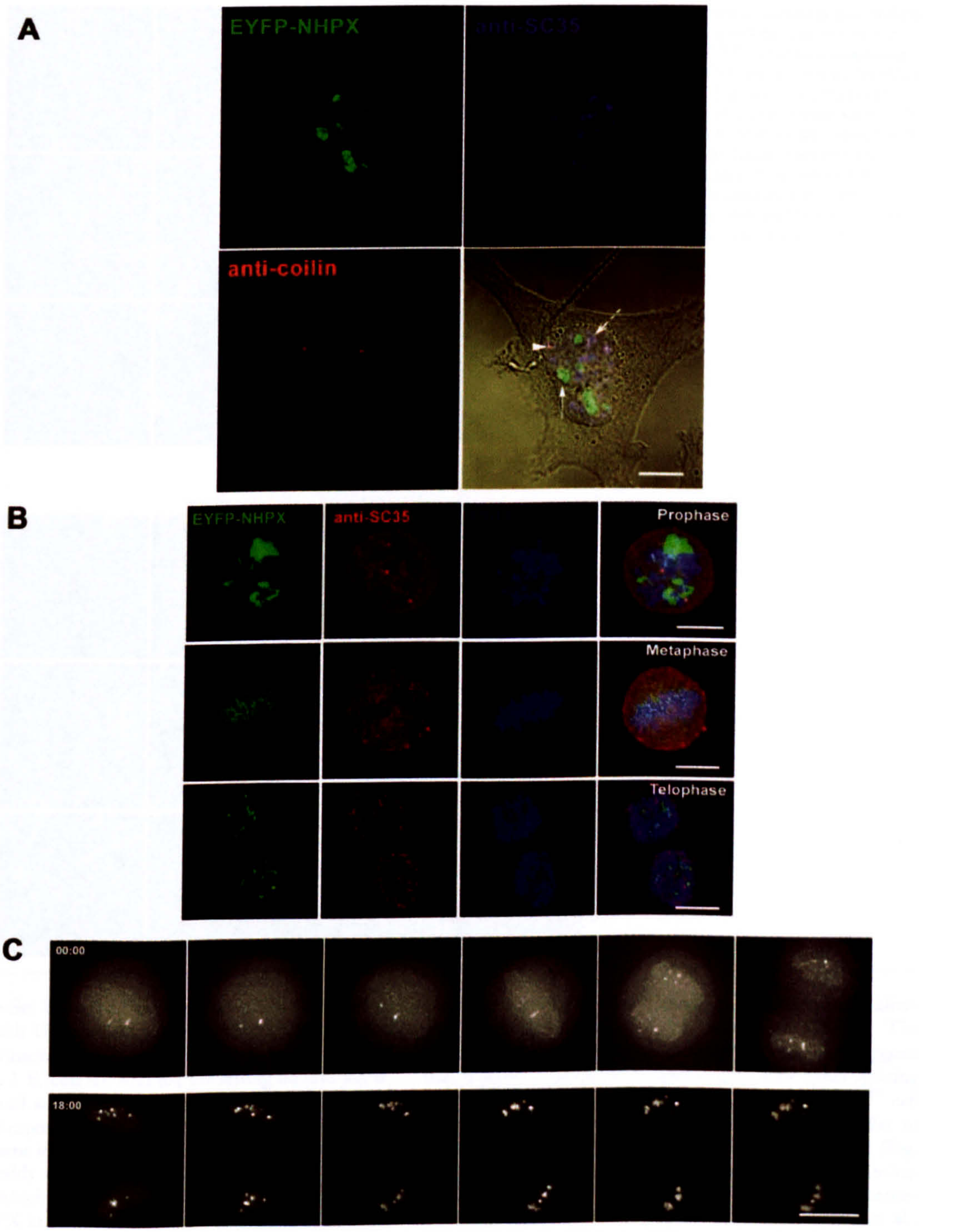


Figure 2. **Localization of NHPX during interphase and mitosis.** The patterns of EYFP–NHPX in HeLa$^{EYFP-NHPX}$ were analyzed (A) during interphase and (B) in different stages of mitosis. Anti-coilin 204/10 was used to denote CBs, anti-SC35 to denote speckles and DAPI to show the condensed chromosome in mitotic stages. Arrowhead indicates CB, arrow indicates nucleolus, and broken arrow indicates speckles. (C) The pattern of EYFP–NHPX in HeLa$^{EYFP-NHPX}$ during mitosis was followed after metaphase in a live cell imaged every 3 min. Bars, 5 μm.

sion protein in vivo is specifically complexed with the same RNA targets that NHPX was shown to bind directly in vitro (Nottrott et al., 1999; Watkins et al., 2000). Given that U4 and U6 snRNAs usually exist as a duplex inside the nucleus, the immunoprecipitation of U4 snRNA alone is surprising. The disruption of the U4/U6 interaction in this assay is unlikely,

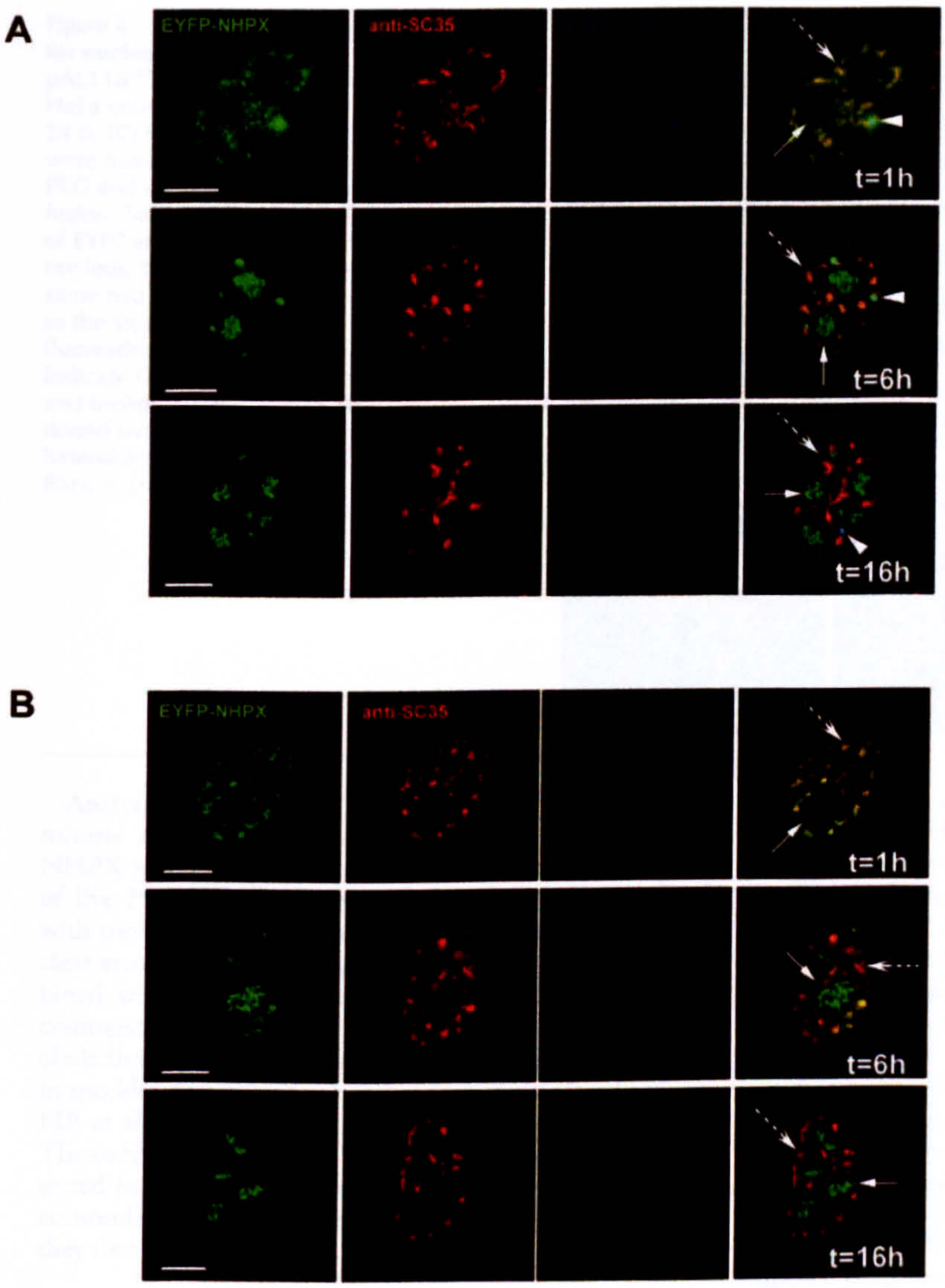


Figure 3. **NHPX transiently accumulates in splicing speckles.** Microinjection of pAL107$^{EYFP-NHPX}$ into (A) transformed cell line MCF7 and (B) primary fibroblast htert1787. The microinjected cells were fixed at different time points (1 h, 6 h and 16 h) and counterstained with anti-SC35 to denote speckles and anti-coilin 204/10 to denote CBs. Arrowheads indicate CBs, arrows indicate nucleoli and broken arrows indicate speckles. Bars, 5 μm.

because a non-Sm RNA, i.e., U6 snRNA, can be coimmunoprecipitated with U4 from the ECFP–SmB cell extract under the same experimental conditions and using the same anti-GFP antibody (Fig. 1 E, lane 4). Another interesting observation is that only a small amount of the total U4 snRNA was pulled down in these experiments (Fig. 1 E, lane 3), even though U4 snRNA is present in similar abundance to U3 snoRNA inside HeLa cells (Reddy and Bush, 1988). This suggests that only a subset of U4, which is likely not bound to U6, may be interacting with NHPX in vivo. Alternatively, the fraction of NHPX bound to the U4/U6 duplex could be inaccessible to the anti-GFP antibodies.

NHPX is colocalized with U3-containing snoRNPs, rather than U4-enriched splicing speckles, during interphase

The ability of NHPX to bind both U3 snoRNA and U4 snRNA, which normally are located in different subnuclear structures, prompted us to investigate further the localization of NHPX under different conditions (Fig. 2). The binding of NHPX to the spliceosomal U4 snRNA suggests that it should colocalize, at least in part, with other splicing factors. However, EYFP–NHPX in the HeLa$^{EYFP-NHPX}$ cell line does not show a speckled nuclear pattern similar to other human splicing factors such as SC35 and U1A (Fig. 2 A, arrowhead indicates CB, arrow indicates nucleolus, broken arrow indicates speckles). The anti-NHPX antiserum also does not label speckles (Fig. 1 B, Chang et al., 1999). Instead, EYFP–NHPX is accumulated in nucleoli and CBs and colocalizes with the snoRNP protein FIB (Figs. 1 A and 2 A ; unpublished data). Also, nucleolar segregation, caused by treating the cells with the transcription inhibitor actinomycin D, results in the relocation of NHPX to the nucleolar periphery along with FIB, but does not cause it to colocalize with either SC35, or other splicing factors (unpublished data).

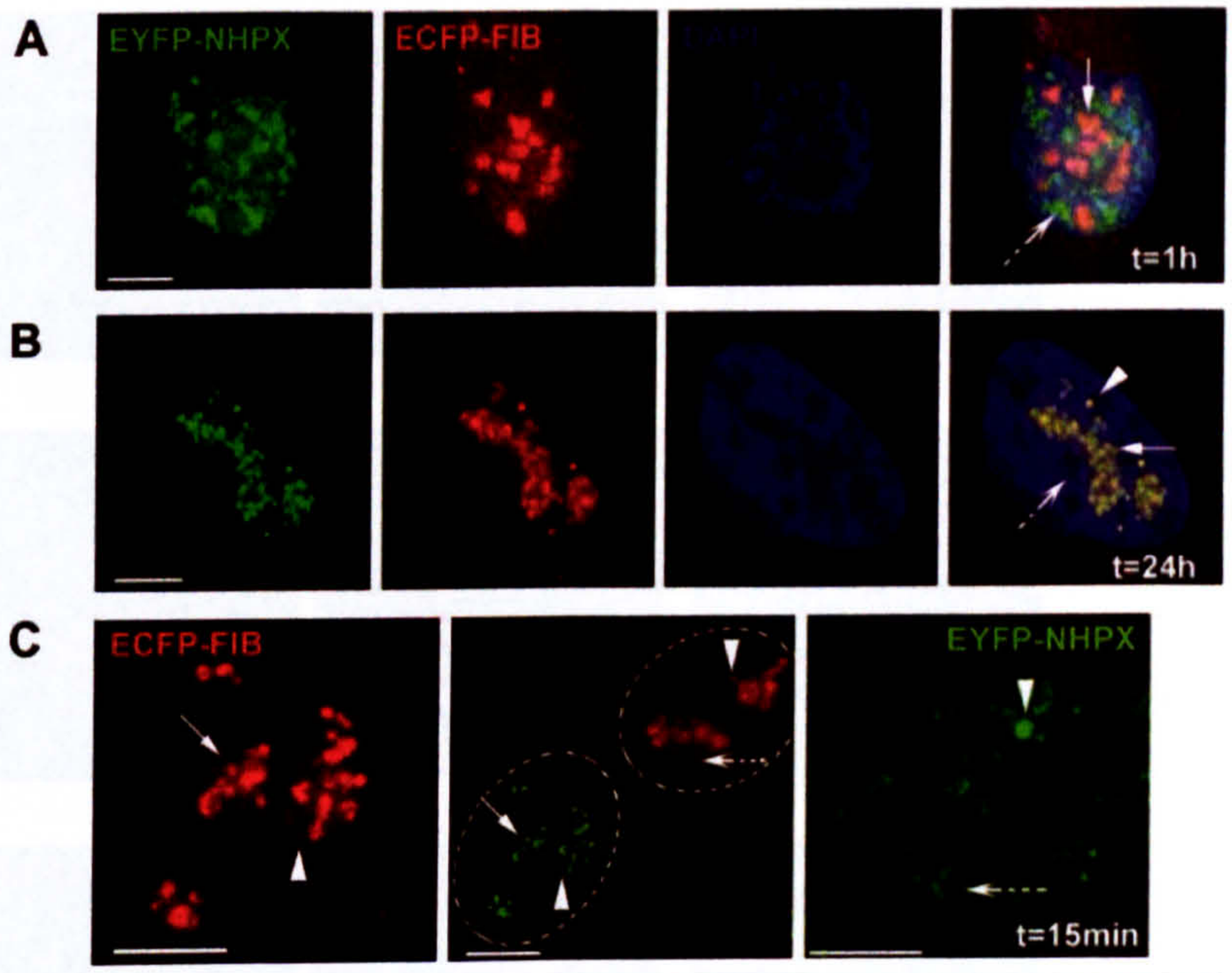


Figure 4. **Separate targeting pathways for nuclear factors.** pAL107$^{EYFP-NHPX}$ and pAL118$^{ECFP-FIB}$ were cotransfected into HeLa cells and fixed after (A) 1 and (B) 24 h. (C) HeLa$^{EYFP-NHPX}$ and HeLa$^{ECFP-FIB}$ were fused to form heterokaryon using PEG and were fixed at 15 min after fusion. To show the relative distribution of EYFP and ECFP components in each nucleus, the central panel shows the same two nuclei within a heterokaryon as the side panels, but in the opposite fluorescence channels. Arrowheads indicate CBs, arrows indicate nucleoli, and broken arrows indicate speckles; dotted ovals outline nuclei in the heterokaryon in the central panel. Bars, 5 μm.

Analysis of the HeLa$^{EYFP-NHPX}$ cells at different stages of mitosis also showed no evidence for the association of NHPX with splicing factors (Fig. 2 B). Time-lapse analysis of live HeLa$^{EYFP-NHPX}$ cells showed that NHPX associated with snoRNP-containing nucleoli immediately after the nuclear envelope reforms (Fig. 2 C; unpublished data). Combined with other immunofluorescence data of fixed cells counterstained with anti-FIB (unpublished data), we conclude that EYFP–NHPX does not accumulate with snRNPs in speckles, but instead colocalizes with the snoRNP protein FIB at all cell cycle stages and metabolic conditions tested. The only colocalization of NHPX with splicing snRNPs detected in vivo was specifically in CBs, structures known to accumulate newly imported snRNPs and snoRNPs when they first enter the nucleus.

Newly imported NHPX transiently colocalizes with splicing factors

Next, we investigated the localization of newly synthesized NHPX because recent data have shown a temporal pathway for snRNP and snoRNP localization in nuclei (Fig. 3; Carvalho et al., 1999; Narayanan et al., 1999a, 1999b; Sleeman and Lamond, 1999a; Sleeman et al., 2001). To our surprise, microinjection of the expression vector pAL107$^{EYFP-NHPX}$ into the cultured cell lines, followed by examination in the fluorescence microscope, revealed that 1 h after microinjection, EYFP–NHPX is accumulated in splicing speckles and CBs (Fig. 3). 2–6 h postmicroinjection, EYFP–NHPX is also detected in nucleoli, whereas the level of EYFP–NHPX in speckles shows a concomitant decrease. At later time points, the signal in speckles can no longer be detected and EYFP–NHPX accumulates specifically in nucleoli and CBs, resembling the pattern observed in the HeLa$^{EYFP-NHPX}$ cell line during interphase.

This novel nuclear pathway for NHPX was observed not only in parental HeLa cells, but also when newly synthesized EYFP–NHPX was expressed after either microinjection or transient transfection in other transformed cell lines, including MCF7 (Fig. 3 A, arrowhead indicates CB, arrow indicates nucleolus, broken arrow indicates speckles) and HEK293 and in primary cell lines, i.e., human foreskin fibroblasts and primary human fibroblasts expressing telomerase htert1787 (Fig. 3 B; unpublished data). Some cell lines, i.e., htert1787, do not contain prominent CBs, and therefore provide an opportunity to test whether the presence of CBs is required for the transport and/or localization of NHPX in nucleoli. Microinjection of pAL107$^{EYFP-NHPX}$ into htert1787 cells did not induce the formation of CBs, and EYFP–NHPX still relocated in the same temporal sequence from speckles to nucleoli (Fig. 3 B, arrow indicates nucleolus, broken arrow indicates speckles). Therefore, the presence of CBs is apparently not required for the directional movement of NHPX from speckles to nucleoli.

Newly synthesized NHPX does not colocalize with U3 snoRNP in speckles

Next, we examined whether the speckle localization of newly synthesized NHPX is a consequence of a previously unknown behavior of U3 snoRNP. To test this, an expression vector, pAL118$^{ECFP-FIB}$, encoding ECFP-tagged FIB, a known U3 snoRNP component, was cotransfected with pAL107$^{EYFP-NHPX}$ into the parental HeLa cell line and analyzed at various time points posttransfection (Fig. 4). At 1 h posttransfection, ECFP–FIB had already accumulated in nucleoli and CBs, whereas in the same cells, EYFP–NHPX accumulated in speckles and CBs, but not in nucleoli (Fig. 4 A, arrow indicates nucleolus, broken arrow indicates speckles). Several hours later, EYFP–NHPX began to accumulate in nucleoli and the signal in speckles decreased, whereas the

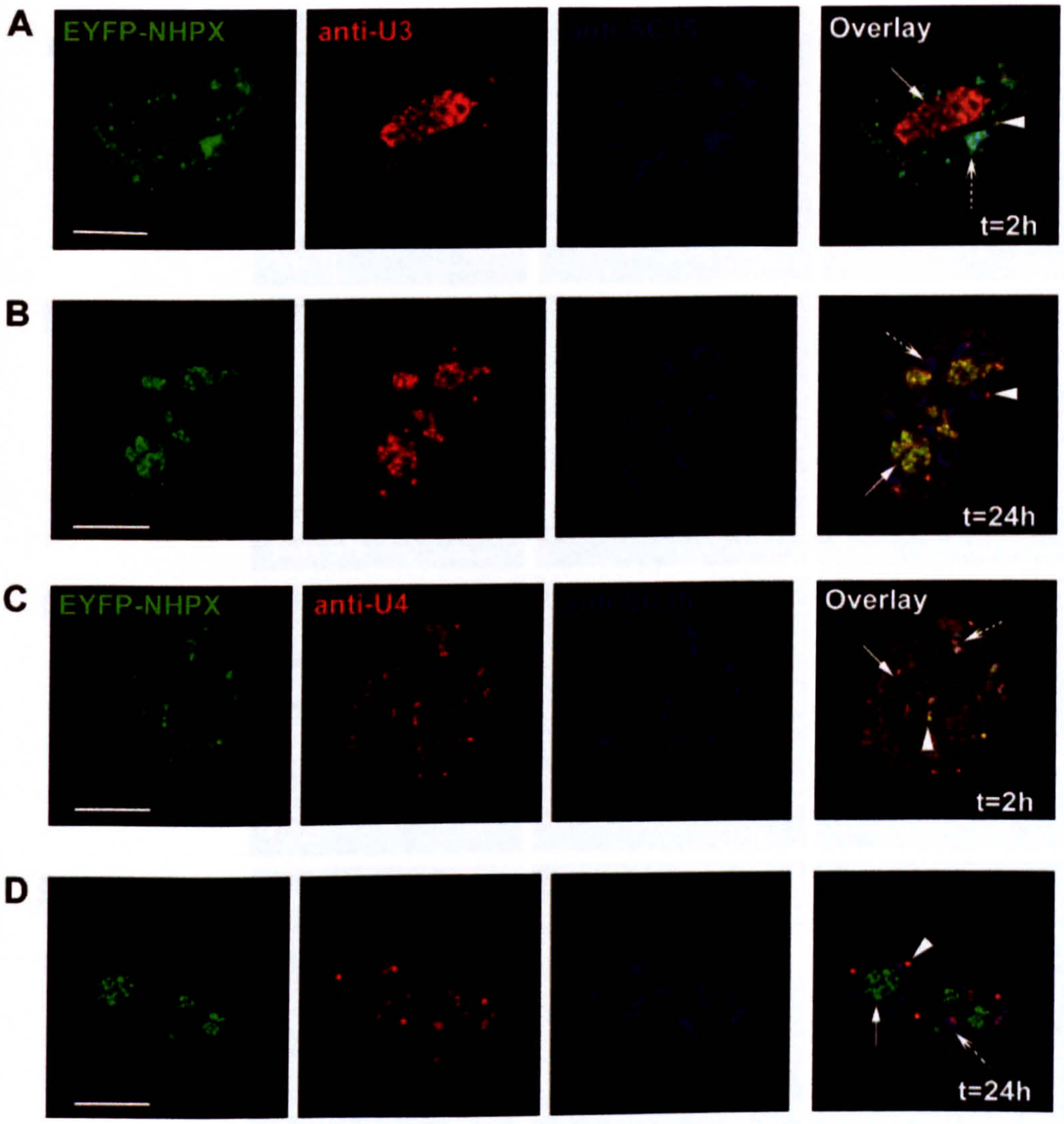


Figure 5. **Localization of endogenous snRNPs and snoRNPs.** pAL107$^{EYFP-NHPX}$ was microinjected into HeLa cells and fixed after (A and C) 2 and (B and D) 24 h. The cells were then counterstained with antisense 2′-*O*-methyl RNA (A and B) U3 and (C and D) U4 and anti-SC35 to show the location of speckles. Arrowheads indicate CBs, arrows indicate nucleoli, and broken arrows indicate speckles. Bars, 5 μm.

ECFP–FIB remained specifically in nucleoli, and both proteins were detected in CBs (unpublished data). After one cell cycle, EYFP–NHPX and ECFP–FIB both quantitatively colocalized in the dense fibrillar component inside nucleoli and in CBs (Figs. 1 A and 4 B, arrowhead indicates CB, arrow indicates nucleolus, broken arrow indicates speckles; unpublished data). Therefore, the transient presence of newly expressed NHPX in nuclear speckles is not explained by its association with U3 snoRNP.

The differential timing in the nucleolar entry of NHPX and FIB is confirmed by analysis of heterokaryons formed between HeLa$^{ECFP-FIB}$ and HeLa$^{EYFP-NHPX}$ stable cell lines (Fig. 4 C). Formation of heterokaryons between two cell lines expressing FP-tagged proteins allows the gradual introduction of EYFP–NHPX into the HeLa$^{ECFP-FIB}$ cells and vice versa (Sleeman et al., 2001). The advantages of using this heterokaryon approach are that it minimizes the possible effect of overexpression that can occur in both microinjection and transient transfection assays, and allows the dynamic exchange of two fluorophores to be observed under the same conditions. ECFP–FIB accumulated directly in nucleoli as soon as 15 min after fusion, whereas EYFP–NHPX accumulated instead in speckles at the same time (Fig. 4 C, arrowheads indicate CBs, arrows indicate nucleoli, broken arrows indicate speckles; unpublished data). Therefore, the NHPX protein detected in speckles is unlikely to be associated with U3 snoRNP.

The expression of exogenous NHPX raises the possibility that the target RNAs it binds to may be upregulated, and thereby results in the observed temporal sequence of nuclear localization. Therefore, we investigated the in vivo localizations of U3 and U4 RNAs in the HeLa cells that were mi-

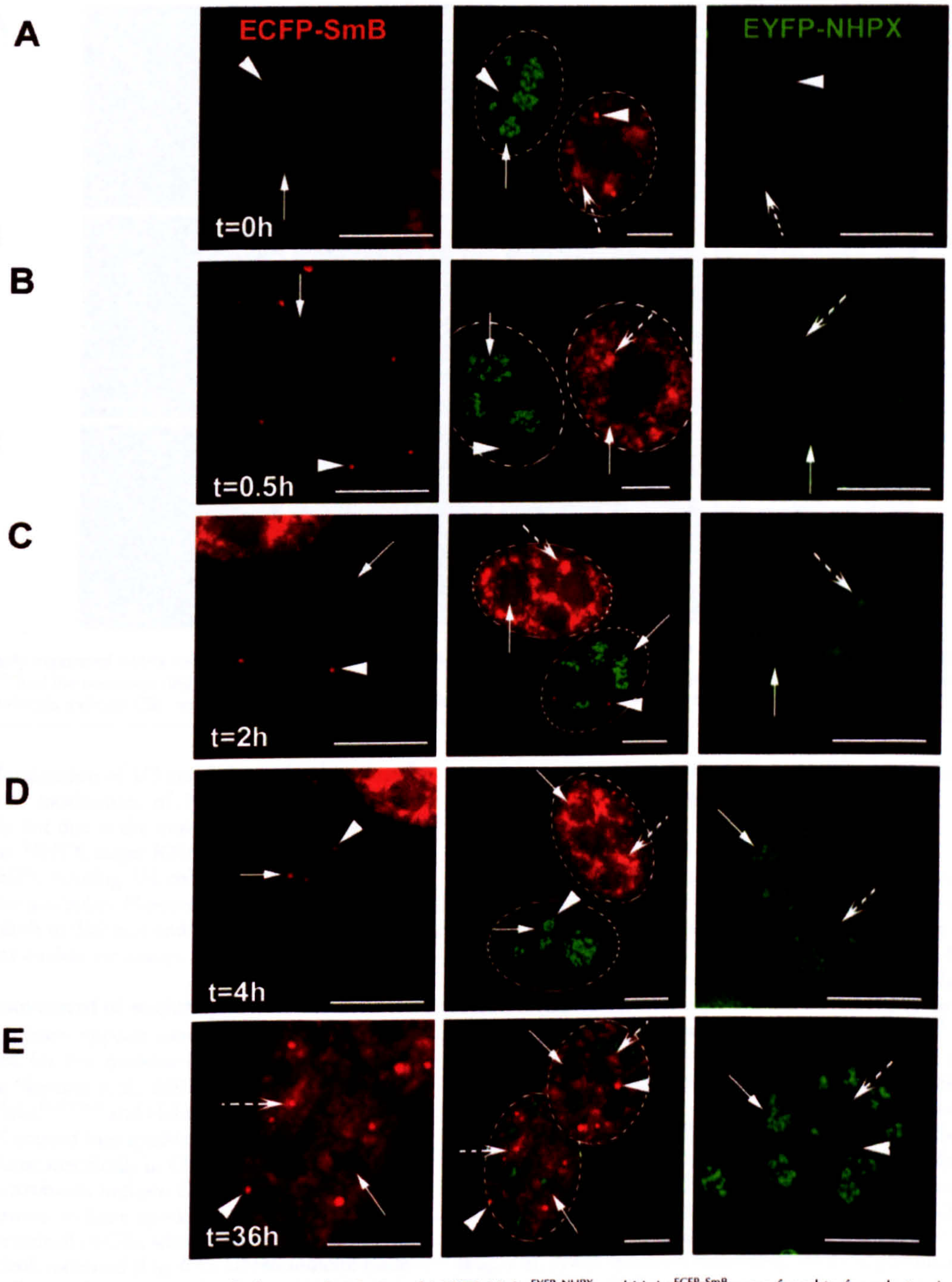


Figure 6. **Reciprocal movement of nucleolar proteins SmB and NHPX.** HeLa$^{EYFP-NHPX}$ and HeLa$^{ECFP-SmB}$ were fused to form heterokaryon using PEG and were fixed at different time points: 0 h (A); 0.5 h (B); 2 h (C); 4 h (D); and 36 h (E). Panel representation as of Fig. 4 C. Arrowheads indicate CBs, arrows indicate nucleoli, and broken arrows indicate speckles; dotted ovals outline nuclei of the heterokaryon in the central panel. Bars, 5 μm.

croinjected with pAL107$^{EYFP-NHPX}$ using complementary 2′-*O*-methyl antisense RNAs (Fig. 5; for review see Lamond, 1993). The U3 in microinjected cells remained localized in the dense fibrillar component of nucleoli and CBs at both early and late time points, similar to control, nonmicroinjected cells (Fig. 5 A and B, arrowhead indicates CB, arrow indicates nucleolus, broken arrow indicates speckles; unpublished data). The U4 snRNA was detected in speckles and CBs in both the microinjected cells at different time points and in control, nonmicroinjected cells (Fig. 5, C and D).

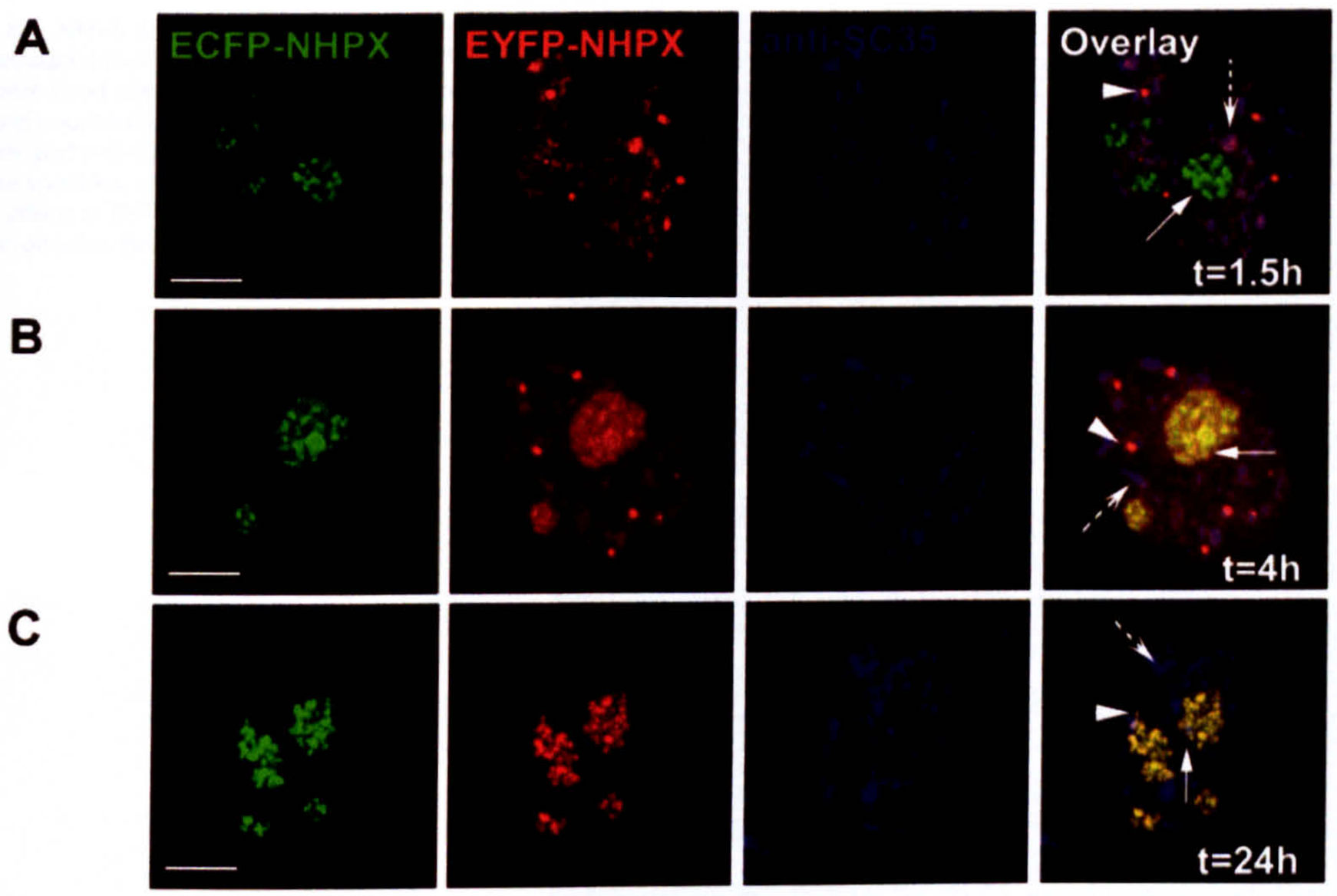


Figure 7. **Newly expressed NHPX localizes to speckles.** HeLa cells were transfected with pAL214$^{ECFP-NHPX}$ for 24 h before microinjecting pAL107$^{EYFP-NHPX}$ and the microinjected cells were fixed after 1.5 h (A), 4 h (B), and 24 h (C), and counterstained with anti-SC35 to denote speckles. Arrowheads indicate CBs, arrows indicate nucleoli, and broken arrows indicate speckles. Bars, 5 μm.

Because the localization of U3 and U4 remained unaltered, the differential localization of NHPX at different time points is likely not due to the movement or relocalization of either of these NHPX target RNAs. These data are consistent with NHPX binding U4 snRNA in speckles and U3 snoRNA in the nucleolus. However, we cannot exclude that NHPX also binds to different and/or unknown target RNAs at the different nuclear structures.

Reciprocal movement of nuclear proteins

The NHPX pathway appears complementary to that previously described for Sm proteins (Fig. 6; Sleeman and Lamond, 1999a; Sleeman et al., 2001). By making heterokaryons between HeLa$^{ECFP-SmB}$ and HeLa$^{EYFP-NHPX}$, we observed that EYFP–NHPX entered into speckles directly, whereas ECFP–SmB accumulated specifically in CBs shortly after fusion (Fig. 6, A and B; arrowheads indicate CBs, arrows indicate nucleoli, broken arrows indicate speckles). At 2 h after fusion, ECFP–SmB remained in CBs, whereas the EYFP–NHPX signal inside nucleoli increased (Fig. 6 C, arrows indicate nucleoli). In some cells, ECFP–SmB was also detected inside nucleoli, as previously reported (Fig. 6 D; Sleeman and Lamond, 1999a; Sleeman et al., 2001). At later time points (~36 h), the ECFP–SmB signal in speckles increased whereas the EYFP–NHPX signal in speckles decreased to a very low/undetectable level (Fig. 6 E; unpublished data). Therefore, we conclude that both nuclear pathways, though operating in different directions, function simultaneously inside a single cell nucleus. The pathways can also be observed by live cell imaging over a period of 12 h (unpublished data). This demonstrates the directed movement of proteins between separate, membrane free nuclear compartments.

Newly expressed NHPX localizes to speckles

We next investigated whether the directed movement of NHPX is either restricted to newly synthesized proteins, or whether it is a reversible relocation of existing proteins (Fig. 7). To test this, pAL214$^{ECFP-NHPX}$ was transfected into HeLa cells and left for 24 h, such that ECFP–NHPX was already accumulated in nucleoli and CBs before microinjection of pAL107$^{EYFP-NHPX}$. Microinjection of pAL107$^{EYFP-NHPX}$ provided a pulse of newly synthesized EYFP–NHPX that accumulated in splicing speckles and CBs, whereas the previously expressed ECFP–NHPX accumulated instead in nucleoli and CBs (Fig. 7 A; arrowhead indicates CB, arrow indicates nucleolus and broken arrow indicates speckles). Gradually, EYFP–NHPX appeared in nucleoli, whereas the signal in speckles subsided. The nucleolar pattern of ECFP–NHPX remained unaltered after microinjection of pAL107$^{EYFP-NHPX}$ (Fig. 7 B, arrow indicates nucleolus). At 24 h postmicroinjection, EYFP–NHPX completely colocalized with the existing ECFP–NHPX (Fig. 7 C, arrowhead indicates CB, arrow indicates nucleolus). These data indicate that only newly synthesized NHPX accumulates in splicing speckles, and further argue that this association is transient. Thus, the presence of NHPX in speckles is likely not a result of protein relocation due to exogenous expression.

To test further whether or not pools of NHPX in speckles and nucleoli interchange, we generated micronuclei by treating the HeLa$^{EYFP-NHPX}$ cells with the spindle-disrupting drug

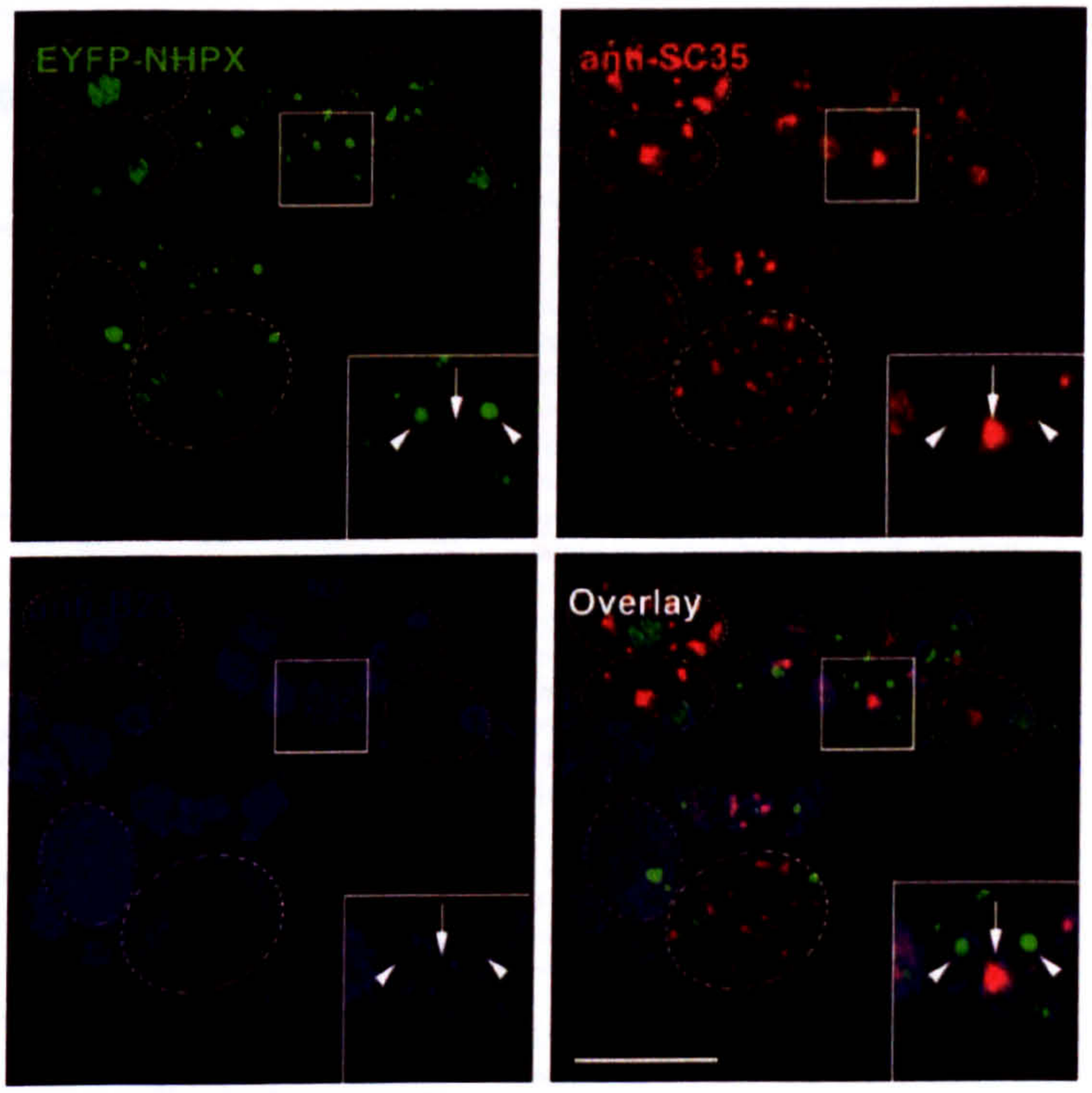


Figure 8. **NHPX does not accumulate in speckles in micronuclei lacking nucleoli.** (A) HeLa$^{EYFP-NHPX}$ cells were fixed after treating with colcemid for 31 h and counterstained with anti-SC35 to denote speckles and anti-B23 to denote nucleoli. Arrows indicate speckles, whereas arrowheads indicate the locations of EYFP–NHPX; dotted ovals outline the micronuclei. Bar, 5 μm.

colcemid (Fig. 8). Colcemid inhibits the progress of mitosis and renders the segregation of chromosomes into many micronuclei without preventing DNA replication, mRNA transcription, splicing, and protein translation (Ferreira et al., 1997). Nucleoli are assembled only on the nucleolar organizer regions in 5 out of 23 chromosomes in human nuclei (Introduction), and therefore the colcemid treatment allows the generation of a subset of micronuclei without nucleoli. To locate those micronuclei, we screened with an antibody specific for nucleolar antigen B23. If the two pools of NHPX in splicing speckles and nucleoli freely exchange, NHPX originally from nucleoli would be expected by default to accumulate back in splicing speckles in the micronuclei lacking nucleoli. Interestingly, EYFP–NHPX does not accumulate in splicing speckles, even in those micronuclei lacking nucleoli (Fig. 8, arrows indicate speckles, arrowheads indicate NHPX localizations, inset shows a micronucleus that lacks nucleoli); instead, they are localized in spot-like structures that also contain the snoRNP protein FIB, but not the CB marker coilin (unpublished data). Microinjection of pAL107$^{EYFP-NHPX}$ into colcemid-treated parental HeLa cells showed the same temporal sequence of localization in speckles prior to nucleoli as seen for untreated cells (unpublished data). Therefore, this differential localization is apparently not a result of colcemid modifying the NHPX pathway and the pools of NHPX localized in splicing speckles and nucleoli appear not to interchange.

The NHPX pathway is unidirectional

The noncycling behavior of NHPX between speckles and nucleoli prompted us to further investigate the directionality of the localization pathway. We performed fluorescence loss in photobleaching (FLIP) analyses of different nuclear structures in the HeLa$^{EYFP-NHPX}$ cells in which one area of the cell is repeatedly bleached while collecting images of the entire cell (Fig. 9). If fluorescent molecules from other regions of the cell diffuse into the bleached area (Fig. 9, white circle indicates bleach zone), loss of fluorescence will occur from both places, indicating that the regions are connected (for review see Reits and Neefjes, 2001). First, we tested whether NHPX inside speckles is moving into the nucleolus (Fig. 9 A, a and c). The positions of speckles were defined by DsRED-U1A in live cells (Fig. 9 A, b and d). The fluorescence intensity of EYFP–NHPX in speckles outside the bleached region decreased, indicating that NHPX diffuses between these nuclear domains (Fig. 9 A, curve b). In comparison, the signals inside nucleoli only showed a minor decrease (Fig. 9 A, curve a). This is consistent with the expected movement of NHPX from speckles to nucleoli. The small change in nucleolar fluorescence may be because the directed movement of NHPX from speckles to nucleoli is slow (hours) compared with the experimental time (~15 min), and/or because it accounts only for a small fraction of the total NHPX signal in nucleoli. However, repeated photobleaching of the nucleolus (Fig. 9 B, white circle indicates bleach zone) resulted in the immediate loss of signal in neighboring nucleoli, indicating that the nucleolar pool of NHPX can cycle between different nucleoli (Fig. 9 B, curve e). The constant level of fluorescence observed in the speckles in the same experiment further strengthens the argument in favor of a unidirectional movement of NHPX from speckles to nucleoli (Fig. 9 B, curve f).

The difference in fluorescence intensity between speckles and nucleoli raises the possibility that the flow from

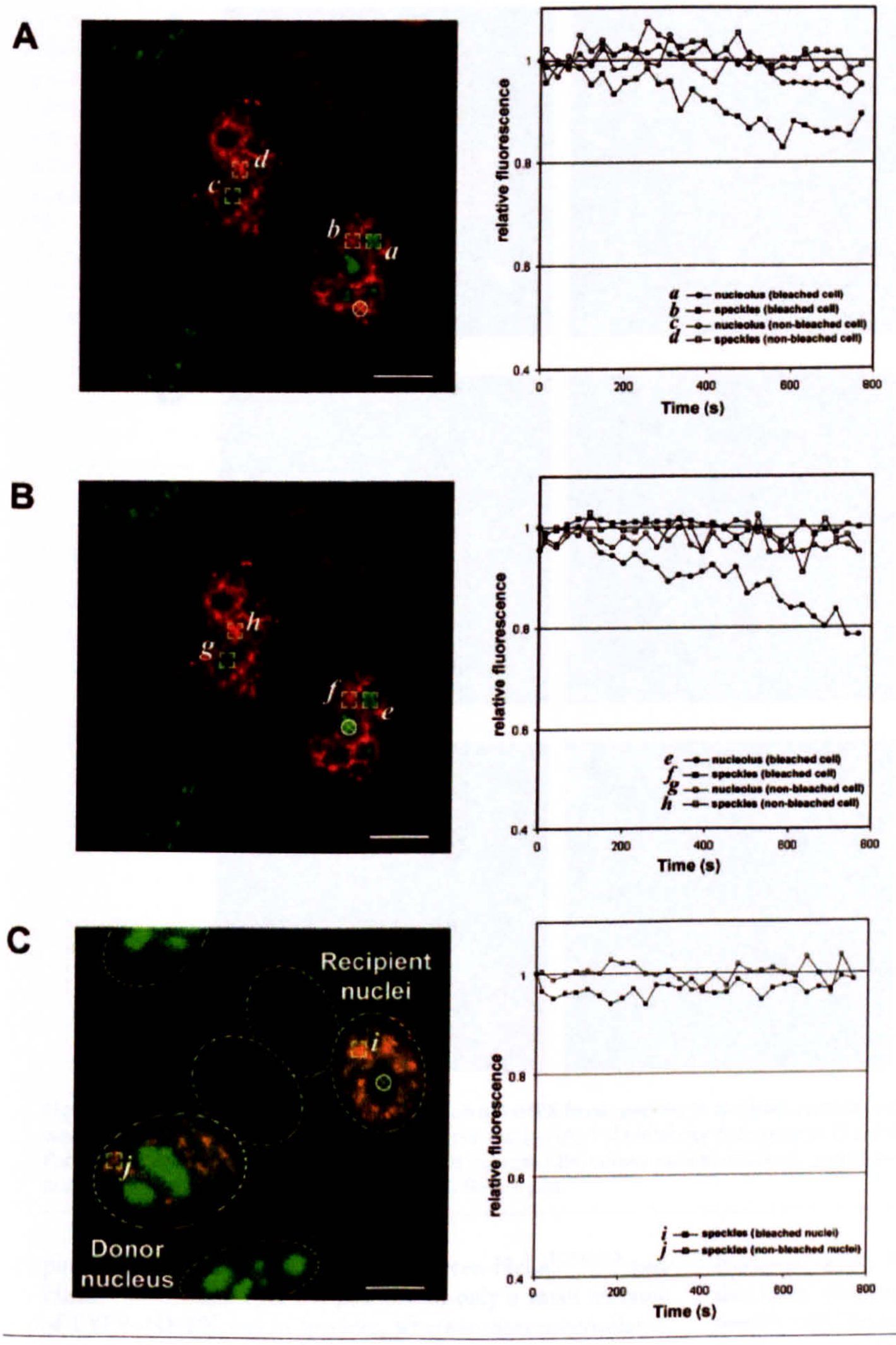


Figure 9. **FLIP analysis of HeLa$^{EYFP–NHPX}$.** A region in the (A) speckles and (B) nucleolus was photobleached repetitively every 20 s and the fluorescence intensities of EYFP–NHPX were analyzed over 15 min. The positions of speckles were located in the live cells using DsRED-U1A that were transfected into the cell lines for 24 h before photobleaching and the selected region for photobleaching were highlighted by the white circle in left panel. The fluorescence intensities of EYFP–NHPX in different regions of the bleached and nonbleached cells were compared in right panel. (C) FLIP analysis of the newly imported EYFP–NHPX in speckles of the heterokaryon formed between HeLa$^{EYFP–NHPX}$ and HeLa cells that were both transfected with pDsRED-U1A for 24 h. The position of the nucleolus for photobleaching (left panel, white circle) in the recipient nuclei were located by both phase contrast microscopy and the absence of DsRED-U1A. The fluorescent intensities of EYFP-NHPX in speckles of bleached and nonbleached nuclei of the heterokaryon were analyzed and shown on the right panel. Dotted ovals outline nuclei in the heterokaryon. Bars, 5 μm.

nucleoli to speckles was not observable in the HeLa$^{EYFP–NHPX}$ cells. To address this, we performed FLIP analysis on heterokaryons formed between HeLa$^{EYFP–NHPX}$ cells and the parental HeLa cells (Fig. 9 C). The locations of nucleoli in the recipient nuclei were positioned both by phase contrast microscopy and by the absence of splicing factor U1A. Shortly after the fusion, as shown previously (Figs. 4 C and 6, A–E), EYFP–NHPX first appeared in splicing speckles but was absent from nucleoli in the recipient nuclei. Repeated photobleaching inside the nucleoli did not change the fluorescence level of EYFP–NHPX inside the speckles of the recipient nuclei (Fig. 9 C, curve i, white circle indicates bleach zone). Therefore, NHPX either does not cycle from nucleoli to speckles, or else does so at a rate/level that cannot be detected in this assay. In summary, the photobleaching analyses, combined with the other data presented here, suggest a unidirectional movement of NHPX from splicing speckles to nucleoli. However, pools of EYFP–NHPX appear freely diffusible between separate components of the same nuclear structure, indicating the regulated entry of nuclear proteins into different domains inside the nucleus.

The progression of NHPX from speckles to nucleoli is dependent on Pol II, but not Pol I transcription

Next, we tested whether the progression of NHPX from speckles to nucleoli requires gene expression, including both Pol I and Pol II transcription (Fig. 10). We again em-

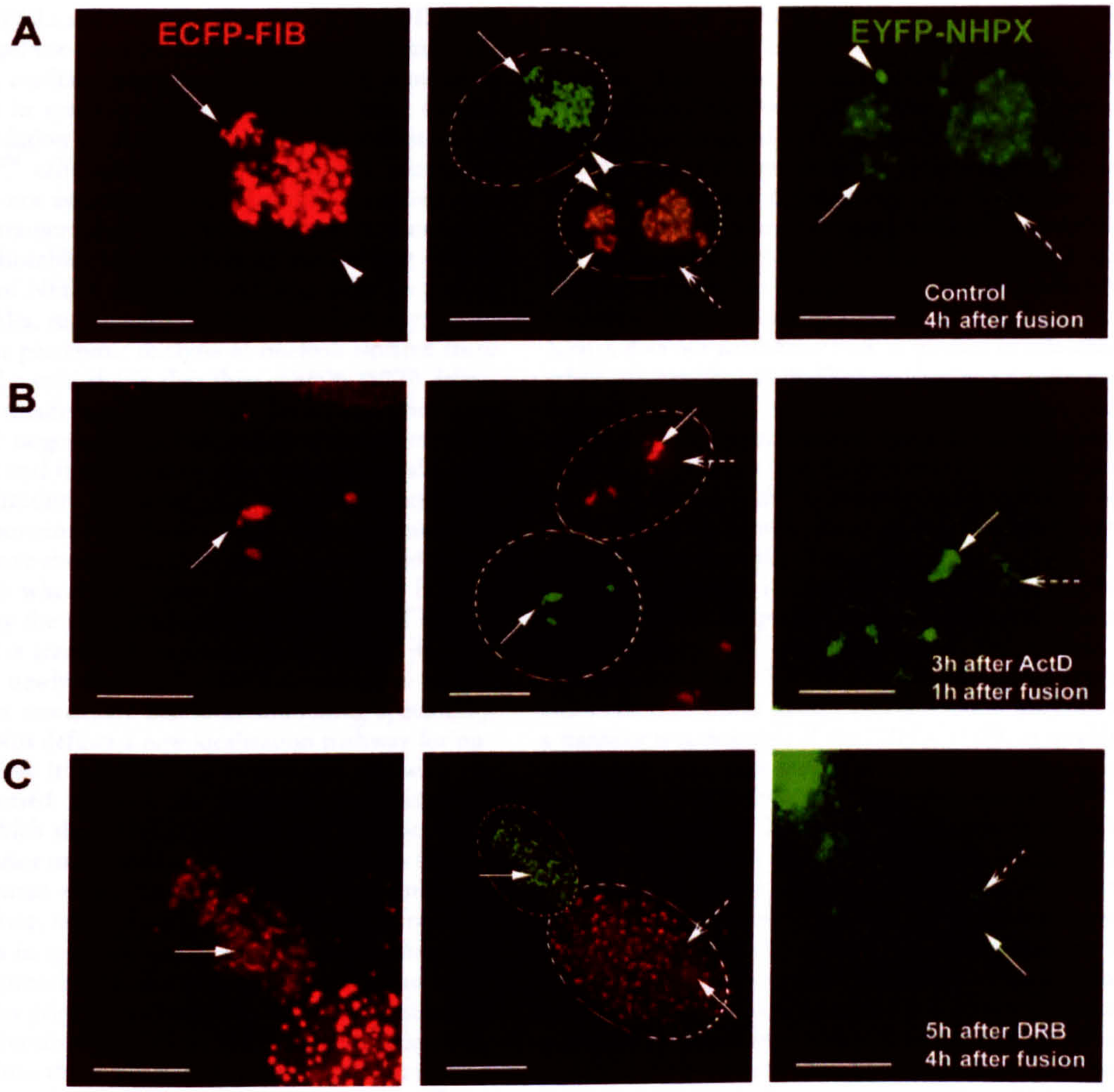


Figure 10. **Transcription-dependent relocation of NHPX from speckles to nucleoli.** Heterokaryon formed between HeLa$^{EYFP-NHPX}$ and HeLa$^{ECFP-FIB}$ were treated with different transcription inhibitors: control (A); Pol I inhibitor Actinomycin D (0.04 μg/ml) (B); and Pol II inhibitor DRB (100 μM) (C). Panel representation as of Fig. 4 C. Arrowheads indicate CBs, arrows indicate nucleoli, and broken arrows indicate speckles; dotted ovals outline nuclei of the heterokaryon in the central panel. Bars, 5 μm.

ployed the heterokaryon approach between HeLa$^{ECFP-FIB}$ and HeLa$^{EYFP-NHPX}$ cells. At 4 h postfusion, only a small amount of EYFP–NHPX was in speckles, whereas most accumulated in nucleoli (Figs. 6 D and 10 A, arrowheads indicate CBs, arrows indicate nucleoli, broken arrows indicate speckles). The heterokaryons formed between HeLa$^{EYFP-NHPX}$ and HeLa$^{ECFP-FIB}$ cells were subjected to transcription inhibitors targeted to specific polymerases. Low levels of actinomycin D cause the segregation of nucleoli and inhibit rRNA transcription, but not pre-mRNA transcription. Newly synthesized EYFP–NHPX moved to the speckles of recipient HeLa$^{ECFP-FIB}$ nuclei, prior to accumulating in the segregated nucleoli (Fig. 10 B, arrows indicate segregated nucleoli, broken arrows indicate speckles), suggesting that pol I transcription and/or ribosome biogenesis is not a prerequisite for the NHPX pathway. Similarly, the immunosuppressant rapamycin, which inhibits transcription of a subset of ribosomal protein genes and hence ribosome assembly, gave the same results (unpublished data). However, when RNA pol II transcription was inhibited, either by α-amanitin or DRB, progression of the newly synthesized NHPX from speckles to the nucleolus was blocked (Fig. 10 C, arrows indicate nucleoli, broken arrows indicate speckles; unpublished data). This suggests that one or more factors must be continually synthesized to allow the newly imported NHPX to move from speckles to nucleoli.

Discussion

In this study we have identified a novel nuclear pathway that leads to the nucleolar accumulation of the NHPX protein. The pathway was detected in multiple mammalian cultured cell lines, including both primary and transformed cells. NHPX was analyzed in vivo, fused to either EYFP or ECFP fluorescent protein tags, and the resulting fusion proteins were shown to have similar localization patterns and RNA binding specificities to the endogenous NHPX. A stably

transformed HeLa cell line that expressed EYFP–NHPX was established and used to demonstrate that, upon its initial entry into the nucleus, newly expressed NHPX transiently accumulates in splicing speckles prior to a later, steady-state accumulation in nucleoli. Further characterization of-HeLa$^{EYFP-NHPX}$ cells indicated that the NHPX protein in speckles was not associated with U3 snoRNP and required RNA pol II transcription for efficient relocation to nucleoli. Additional photobleaching experiments showed that the nucleolar pool of NHPX did not interchange with the pool in nuclear speckles, suggesting a unidirectional pathway.

Our recent proteomic analysis of nucleoli isolated from cultured HeLa cells shows that they contain >270 different proteins (Andersen et al., 2002). So far there has been no nucleolar targeting motif identified common to all of these factors and it seems likely that multiple, parallel nucleolar localization pathways can operate. Nonetheless, analyses of proteins that show a steady-state accumulation in nucleoli have shown that they usually move rapidly into the nucleolus when they enter the nucleus. This is illustrated here by the rapid nucleolar accumulation of EYFP–FIB when it is transiently expressed in vivo (Fig. 4). The finding that newly expressed NHPX accumulates in nuclear speckles transiently before accumulating specifically in the nucleolus defines a new localization pathway for nucleolar proteins. It is interesting to compare this with the recently reported pathway for nucleolar localization of snoRNAs, which showed that multiple snoRNAs accumulate in CBs prior to nucleoli upon initial entry into the nucleus (Narayanan et al., 1999a, 1999b; Verheggen et al., 2001). However, none of the snoRNAs showed a transient accumulation in speckles, consistent with our finding that the snoRNP protein FIB also does not transiently accumulate in speckles prior to nucleoli (Fig. 4). We also observed that NHPX localizes to CBs as well as speckles upon its initial entry into the nucleus, but unlike its transient association with speckles, NHPX is also detected in CBs at later stages of expression when the bulk of the protein is concentrated in nucleoli. At present, we cannot distinguish whether NHPX accumulates in CBs prior to speckles, or in both structures at the same time. However, the CB association does not appear to be obligatory for the NHPX localization pathway because a similar transient association with speckles prior to nucleolar accumulation is observed in cell lines lacking prominent CBs (Fig. 3 B). However, similar molecular events may occur either within the nucleoplasm or in CBs that are too small to detect.

It is also interesting to compare the NHPX pathway with the recently identified nuclear pathway for splicing snRNPs where FP-tagged snRNP Sm proteins accumulate in CBs, and nucleoli, prior to speckles, upon their initial nuclear entry (Sleeman and Lamond, 1999a; Sleeman et al., 2001). Therefore, this pathway appears to be complementary to that of NHPX. By analyzing heterokaryons formed between separate stable HeLa cell lines expressing EYFP–NHPX and ECFP–SmB, we could show that both these complementary pathways can operate simultaneously within the same nuclei (Fig. 6). These data confirm the specificity of the pathways and highlight the dynamic mechanisms operating to organize the distribution of proteins and RNPs in the nucleus. The results also point to the localization specificity of the separate subnuclear bodies, including nucleoli, CBs, and speckles, although they are not enclosed by membranes.

In order to answer why NHPX shows the observed transient accumulation in speckles prior to nucleoli, it may be important to consider that it is specifically newly expressed and imported NHPX protein that is detected in speckles. Several experiments showed that NHPX does not localize to speckles by default, and that the nucleolar pool of NHPX does not cycle continually to and from speckles. For example, in micronuclei that lack NOR-containing chromosomes (and hence do not have nucleoli), NHPX does not accumulate back in speckles or colocalize with splicing factors (Fig. 8). FLIP photobleaching experiments also showed that whereas nucleolar EYFP–NHPX can exchange rapidly between separate nucleoli within the same nucleus, little or no exchange occurs with the pool of NHPX in speckles (Fig. 9). This contrasts with the behavior of the nucleolar protein PSP1, which was recently shown to cycle continually between nucleoli and paraspeckles (Fox et al., 2002).

Our data strongly indicate that the association of NHPX with speckles is a temporal phenomenon linked to the entry of newly expressed NHPX into the nucleus. For example, transient expression of EYFP–NHPX in HeLa cells expressing ECFP–NHPX, which already accumulated in nucleoli, shows a transient accumulation of the EYFP–NHPX in speckles before it later colocalizes quantitatively with the existing nucleolar ECFP–NHPX (Fig. 7). Also, when the HeLa$^{EYFP-NHPX}$ cells undergo mitosis, EYFP–NHPX immediately relocalizes to the reforming nucleoli during telophase and does not accumulate in speckles in the postmitotic nuclei (Fig. 2). Therefore, the speckles association is not a result of nuclear import of NHPX per se, but rather relates to an effect specific for newly expressed protein. We propose that a likely explanation for this behavior of NHPX could be related to it having a function required for the assembly or maturation of some form of nuclear protein or RNP complex, prior to its subsequent stable association with U3 and/or other nucleolar snoRNPs. This could imply either that the affinity of NHPX for different target RNAs changes after it enters the nucleus for the first time or that its access to bind snoRNA targets is initially restricted.

Based upon the results of previous biochemical studies on the structure and binding specificity of NHPX, the U4 snRNA is a possible candidate target for NHPX in speckles. NHPX binds U4 snRNA in vitro via the 5′ stem loop sequence (Nottrott et al., 1999; Vidovic et al., 2000). Consistent with this idea, U4 snRNA has been localized to speckles in HeLa cells by hybridization experiments (Carmo-Fonseca et al., 1992; Fig. 5). The fact that we show here that EYFP-NHPX likely interacts in vivo with a form of U4 snRNA that is not stably associated with U6 snRNA suggests that NHPX may transiently interact in speckles with an immature form of U4 snRNP (Fig. 1 E). The observed requirement for gene expression in order for NHPX to move from speckles to nucleoli might reflect a requirement for other factors to be expressed to allow NHPX to complete its transient role in speckles (Fig. 10). Whether this is connected to U4 snRNP assembly and/or some other events remains to be established. Future studies will aim to analyze further the molecular mechanism involved in the novel nucleolar localization pathway detected for NHPX and to establish what biological role this may play.

Materials and methods

Plasmid constructs

NHPX cDNA (gi:26185777) was isolated for PCR amplification from Marathon-Ready HeLa library (CLONTECH Laboratories, Inc.) using specific primers with BglII and KpnI restriction site attached on the 5′ and 3′ primer, respectively. The amplified fragment was subsequently cloned to the BglII-KpnI fragment of EYFP–C1 and ECFP–C1 to form pAL107$^{EYFP-NHPX}$ and pALZ14$^{ECFP-NHPX}$, respectively, and verified by DNA sequencing.

Cell culture, transfection, and establishment of stable cell line

HeLa, MCF7, HEK293, and primary fibroblast htert1787 were grown in DME supplemented with 10% FCS and 100 U/ml penicillin and streptomycin (Invitrogen). 2 μg plasmid construct per 100-mm dish was used for transfection using Effectene (QIAGEN) according to the manufacturer's instruction. EYFP–NHPX and EYFP–FIB stable cell lines were generated essentially as described in Sleeman et al. (2001). ECFP–SmB stable cell line (CFPSmBE8.6) was previously described (Sleeman et al., 2001). EGFP–H2B stable cell line was a gift from T. Kanda (The Salk Institute for Biological Studies, La Jolla, CA) (Kanda et al., 1998). Drug treatments were carried out as follows: Actinomycin D (0.04 μg/ml, 3 h; Sigma-Aldrich); DRB (100 μM, 4 h; Sigma-Aldrich); α-amanitin (40 μM, 4 h; Sigma-Aldrich); and colcemid (0.5 μg/ml, 31 h; Sigma-Aldrich).

FACS analysis

Parental HeLa and HeLa$^{EYFP-NHPX}$ cells were harvested by trypsinization, and fixed in 70% ethanol for 3 h at 4°C. Cells were stained with PI (25 μg/ml) containing RNase A (100 μg/ml). Fluorescence was measured using a FACScan (Becton Dickinson). Cell debris and fixation artifacts were gated out. Data analysis was done using Cell Quest software (Becton Dickinson).

Immunoprecipitation and immunoblotting

For immunoprecipitation, 100 μg protein G Sepharose (Amersham Pharmacia Biotech) was preincubated with 10 μg anti-GFP antibodies (Roche). Extracts were prepared using nuclear lysis buffer (Sleeman et al., 2001). Extracts were precleared with 100 μg protein G Sepharose and then incubated with antibody-bound protein G Sepharose for 16 h. Beads were then washed three times with the lysis buffer, and bound RNAs were released by Proteinase K (2.24 mg/ml) in extraction buffer (0.63% SDS, 26.25 mM EDTA, 26.25 mM Tris-HCl, pH 8.0, 0.28 mg/ml yeast tRNA) for 45 min at 65°C. RNAs were precipitated by adding 7 vol of EtOH/NH_4OAc (86% EtOH, 0.57 M NH_4OAc), and washed once with 70% EtOH. Northern hybridizations were done by standard procedures (Lamond et al., 1989).

For immunoblotting, protein samples were separated by 4–12% Bis-Tris gels (Novex), and subsequently transferred onto nitrocellulose membrane using a submarine system (Novex). After blocking with 5% milk powder in PBS+0.05% Tween-20, the membranes were incubated with either rabbit anti-NHPX R86 (1:100; Chang et al., 1999) or mouse monoclonal anti-GFP (1:1,000; Roche), and the bound antibody was then probed using anti–rabbit HRP (1:2,000; Pierce Chemical Co.) and anti–mouse HRP conjugate, respectively (1:5,000; Pierce Chemical Co.) in PBS containing 5% milk powder and 0.05% Tween-20, and detected via chemiluminescence with ECL Plus (Amersham Pharmacia Biotech).

Immunostaining and 2′-*O*-methyl RNA hybridization

Cells grown on coverslips were washed in PBS and fixed for 10 min in 3.7% (wt/vol) paraformaldehyde in CSK buffer (10 mM Pipes, pH 6.8, 10 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 2 mM EDTA) at RT, permeabilized with 1% Triton X-100 in PBS for 10 min at room temperature, mounted onto glass slides using VectorShield (Vector Lab), and imaged as described below. Antibodies used were anti-FIB monoclonal 72b9 (1:10; Turley et al., 1993); anti-NHPX peptide antibody R86 (1:100; Chang et al., 1999); anti-SC35 monoclonal (1:500; Sigma-Aldrich); anti-coilin 204/10 (1:300; Bohmann et al., 1995); anti-B23 (1:75; Santa Cruz Biotechnology); and TRITC-, Texas red–, and Cy5-conjugated secondary antibodies (Jackson Laboratories). Before being mounted on slides, coverslips were incubated with 1 μM DAPI (Sigma-Aldrich) for 30 s to stain DNA. Fluorescence microscopy of fixed cells was carried out using a 100× NA 1.4 Plan-Apochromat objective. Three-dimensional images and sections were recorded either on a LSM410 Confocal microscope (ZEISS) or on a Zeiss DeltaVision Restoration microscope (Applied Precision, Inc.). Images presented here are maximal intensity projections of the entire nuclear fluorescence.

For 2′-*O*-Methyl RNA hybridization (Carmo-Fonseca et al., 1992), cells were permeabilized with 0.5% Triton X-100 in CSK buffer containing Complete (Roche) on ice for 3 min, and were then fixed in freshly prepared 3.7% paraformaldehyde in CSK buffer for 10 min at room temperature. Cells were washed three times in PBS, one time with 6 × SSPE, and prehybridized with 6 × SSPE/5 × Denhardt's solution containing yeast tRNA (0.5 mg/ml) for 15 min. Cells were then hybridized with the same buffer with biotinylated 2′-*O*-methyl antisense oligonucleotide probe (2 μM) for 30 min, and then were washed three times with 6 × SSPE and rinsed with avidin wash buffer (0.03 M Hepes, pH 7.9, 0.15 M KCl, 0.05% Tween-20, 1% donkey serum) before incubating with Texas red–conjugated avidin DCS (Vector Labs) at 2 μg/ml for 30 min. They were then washed and mounted on slides for microscopic studies as above. (U3 probe 1011: 5′ – *C*CUUUCGGUGCUC*C*C – 3′; U4 probe 1012: 5′ – *C*CUGCCACUGCGCAAAGCU*C*C – 3′; * denotes biotinylated sites).

Mitotic studies of living cells

Cells were grown on 40-mm glass coverslips (Intracel) in medium containing 200 μg/ml G418. Cells were maintained at 37°C by use of a closed-system perfusion chamber (Bioptech) in DME media (20 mM Hepes, without Phenol red; Invitrogen). Images were collected using the 100× NA 1.4 Plan-Apochromat objective on the Zeiss-DeltaVision Restoration microscope. For each nucleus, 20–30 optical sections in the z-axis were recorded. The Hg lamp was attenuated with a 0.5-OD neutral density filter, and images were recorded every 3 min over a time period of 2 h (3 × 3 binning). Time-lapse images were viewed as maximal intensity projections of each time point (SoftWoRx; Applied Precision, Inc.).

Microinjection and heterokaryon formation

For microinjection, pAL107 was diluted to 20 μg/ml with injection buffer (100 mM glutamic acid, pH 7.2 [with citric acid], 140 mM KOH, 1 mM $MgSO_4$, and 1mM DTT) prior to injection into living cells using an Eppendorf 5242 microinjector. For heterokaryon formation, two different cell lines expressing fluorescent proteins were mixed in a ratio of 1:1 in 100-mm diameter petri dishes containing coverslips and cultured until 80–90% confluent. The culture medium was drained, and 1.5 ml of 50% Polyethylene Glycol (PEG hybrimax; Sigma-Aldrich) was added. The dishes were rocked gently for 90 s and washed thoroughly by several changes of fresh culture medium (a modification of Sleeman et al. 2001).

Photobleaching analysis

Cells were grown on 42-mm glass coverslips (no. 1; Helmut Sauer) in medium containing 200 μg/ml G418. Cells were maintained at 37°C by use of a closed perfusion chamber (Bachofer) in DME media (20 mM Hepes, no Phenol red; Invitrogen). Photobleaching experiments were carried out on a Zeiss 510 confocal laser scanning microscope equipped with an argon-krypton laser (ZEISS). The 488-nm laser and a 63× plan Apolens with a 1.4 NA and a laser power of 2.5% was used for image acquisition, and 25% was used for photobleaching. An area of 16 × 16 pixels was bleached with an iteration of 250 (duration of bleach was 3 s). An image was collected after every bleaching event, with 20-s intervals between each bleaching event over a period of 15 min. To locate splicing speckles in vivo, plasmid pDsRED-U1A was transfected into the cells 24–36 h before imaging. Speckles were defined by red fluorescence, whereas nucleoli were defined by both phase contrast and the absence of U1A.

Online supplemental material

Fig. S1 (available at http://www.jcb.org/cgi/content/full/200201120/DC1) depicts immunofluorescence labeling of HeLa cells using anti-NHPX antibodies either with (A) or without (B) transient transfection with pAL107$^{EYFP-NHPX}$. Panel A is identical to Fig. 1 B in the text.

We thank our colleges in the Lamond laboratory, Dr. Laura Trinkle-Mulcahy for plasmid DsRED-U1A, Dr. Judith Sleeman for the ECFP-SmB cell line, Ursula Ryder for help in RNA manipulation, and other colleagues in the laboratory and Prof. C Proud for helpful discussions. We would also like to thank Dr. T. Kanda for the EGFP-H2B cell line, Dr. K. Collins (University of California, Berkeley, Berkeley, CA) for the htert1787 primary fibroblast cell line, and Dr. B. Chen (Dana-Farber Cancer Institute, Boston, MA) for the affinity-purified anti-NHPX rabbit antiserum, Dr. B. Frenguelli for use of the Zeiss LSM 510 confocal microscope, and S. Shreeman for help in FACS analysis.

This work was supported by the Wellcome Trust. A.I. Lamond is a Wellcome Trust Principal Research Fellow. A.K.L. Leung is a Croucher Foundation Scholar (HK) and ORS awardee (UK).

Submitted: 28 January 2002
Revised: 20 March 2002
Accepted: 22 March 2002

References

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. *Curr. Biol.* 12:1–11.

Blobel, G., and V.R. Potter. 1967. Ribosomes in rat liver: an estimate of the percentage of free and membrane-bound ribosomes interacting with messenger RNA in vivo. *J. Mol. Biol.* 28:539–542.

Bohmann, K., J.A. Ferreira, and A.I. Lamond. 1995. Mutational analysis of p80 coilin indicates a functional interaction between coiled bodies and the nucleolus. *J. Cell Biol.* 131:817–831.

Boudonck, K., L. Dolan, and P.J. Shaw. 1999. The movement of coiled bodies visualized in living plant cells by the green fluorescent protein. *Mol. Biol. Cell.* 10:2297–2307.

Carmo-Fonseca, M., L. Mendes-Soares, and I. Campos. 2000. To be or not to be in the nucleolus. *Nat. Cell Biol.* 2:E107–E112.

Carmo-Fonseca, M., R. Pepperkok, M. Carvalho, and A. Lamond. 1992. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J. Cell Biol.* 117:1–14.

Carvalho, T., F. Almeida, A. Calapez, M. Lafarga, M.T. Berciano, and M. Carmo-Fonseca. 1999. The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. *J. Cell Biol.* 147:715–728.

Chang, M.S., H. Sasaki, M.S. Campbell, S.K. Kraeft, R. Sutherland, C.Y. Yang, Y. Liu, D. Auclair, L. Hao, and H. Sonoda, et al. 1999. HRad17 colocalizes with NHP2L1 in the nucleolus and redistributes after UV irradiation. *J. Biol. Chem.* 274:36544–36549.

Dundr, M., T. Misteli, and M.O. Olson. 2000. The dynamics of postmitotic reassembly of the nucleolus. *J. Cell Biol.* 150:433–446.

Ferreira, J., G. Paolella, C. Ramos, and A.I. Lamond. 1997. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J. Cell Biol.* 139:1597–1610.

Fox, A.H., Y.W. Lam, A.K. Leung, C.E. Lyon, J. Andersen, M. Mann, and A.I. Lamond. 2002. Paraspeckles. A novel nuclear domain. *Curr. Biol.* 12:13–25.

Gall, J.G. 2000. Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.* 16: 273–300.

Gall, J.G., M. Bellini, Z. Wu, and C. Murphy. 1999. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol. Biol. Cell.* 10:4385–4402.

Kanda, T., K.F. Sullivan, and G.M. Wahl. 1998. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* 8:377–385.

Lafontaine, D.L., and D. Tollervey. 1998. Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.* 23:383–388.

Lamond, A.I. 1993. 2′-*O*-alkyloligoribonucleotides: probes for studying the biochemistry and cell biology of RNA processing. *Biochem. Soc. Trans.* 21:1–8.

Lamond, A.I., and W.C. Earnshaw. 1998. Structure and function in the nucleus. *Science.* 280:547–553.

Lamond, A.I., B. Sproat, U. Ryder, and J. Hamm. 1989. Probing the structure and function of U2 snRNP with antisense oligonucleotides made of 2′-OMe RNA. *Cell.* 58:383–390.

Lewis, J.D., and D. Tollervey. 2000. Like attracts like: getting RNA processing together in the nucleus. *Science.* 288:1385–1389.

Lyon, C.E., K. Bohmann, J. Sleeman, and A.I. Lamond. 1997. Inhibition of protein dephosphorylation results in the accumulation of splicing snRNPs and coiled bodies within the nucleolus. *Exp. Cell Res.* 230:84–93.

Malatesta, M., C. Zancanaro, T.E. Martin, E.K. Chan, F. Amalric, R. Luhrmann, P. Vogel, and S. Fakan. 1994. Is the coiled body involved in nucleolar functions? *Exp. Cell Res.* 211:415–419.

Matera, A.G. 1999. Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol.* 9:302–309.

Misteli, T. 1999. RNA splicing: What has phosphorylation got to do with it? *Curr. Biol.* 9:R198–R200.

Misteli, T. 2000. Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. *J. Cell Sci.* 113:1841–1849.

Misteli, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. *Science.* 291:843–847.

Muratani, M., D. Gerlich, S.M. Janicki, M. Gebhard, R. Eils, and D.L. Spector. 2001. Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus. *Nat. Cell Biol.* 4:106–110.

Narayanan, A., A. Lukowiak, B.E. Jady, F. Dragon, T. Kiss, R.M. Terns, and M.P. Terns. 1999a. Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J.* 18:5120–5130.

Narayanan, A., W. Speckmann, R. Terns, and M.P. Terns. 1999b. Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol. Biol. Cell.* 10:2131–2147.

Nottrott, S., K. Hartmuth, P. Fabrizio, H. Urlaub, I. Vidovic, R. Ficner, and R. Luhrmann. 1999. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5′ stem-loop of U4 snRNA. *EMBO J.* 18:6119–6133.

Ochs, R.L., T.W. Stein, Jr., and E.M. Tan. 1994. Coiled bodies in the nucleolus of breast cancer cells. *J. Cell Sci.* 107:385–399.

Olson, M.O., M. Dundr, and A. Szebeni. 2000. The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol.* 10:189–196.

Pederson, T. 1998. The plurifunctional nucleolus. *Nucleic Acids Res.* 26:3871–3876.

Phair, R.D., and T. Misteli. 2001. Kinetic modelling approaches to in vivo imaging. *Nat. Rev. Mol. Cell Biol.* 2:898–907.

Platani, M., I. Goldberg, J.R. Swedlow, and A.I. Lamond. 2000. In vivo analysis of Cajal body movement, separation, and joining in live human cells. *J. Cell Biol.* 151:1561–1574.

Reddy, R., and H. Bush. 1988. Small nuclear RNAs: RNA sequences, structure, and modifications. *In* Small Nuclear Ribonucleoprotein Particles. M.L. Birnstiel, editor. Springer-Verlag, New York. 1–37.

Reits, E.A., and J.J. Neefjes. 2001. From fixed to FRAP: measuring protein mobility and activity in living cells. *Nat. Cell Biol.* 3:E145–E147.

Saito, H., T. Fujiwara, S. Shin, K. Okui, and Y. Nakamura. 1996. Cloning and mapping of a human novel cDNA (NHP2L1) that encodes a protein highly homologous to yeast nuclear protein NHP2. *Cytogenet. Cell Genet.* 72:191–193.

Scheer, U., and R. Hock. 1999. Structure and function of the nucleolus. *Curr. Opin. Cell Biol.* 11:385–390.

Schul, W., L. de Jong, and R. van Driel. 1998. Nuclear neighbours: the spatial and functional organization of genes and nuclear domains. *J. Cell. Biochem.* 70: 159–171.

Sleeman, J.E., and A.I. Lamond. 1999a. Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.* 9:1065–1074.

Sleeman, J.E., and A.I. Lamond. 1999b. Nuclear organization of pre-mRNA splicing factors. *Curr. Opin. Cell Biol.* 11:372–377.

Sleeman, J.E., C.E. Lyon, M. Platani, J.P. Kreivi, and A.I. Lamond. 1998. Dynamic interactions between splicing snRNPs, coiled bodies and nucleoli revealed using snRNP protein fusions to the green fluorescent protein. *Exp. Cell Res.* 243:290–304.

Sleeman, J.E., P. Ajuh, and A.I. Lamond. 2001. snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. *J. Cell Sci.* 114:4407–4419.

Snaar, S., K. Wiesmeijer, A.G. Jochemsen, H.J. Tanke, and R.W. Dirks. 2000. Mutational analysis of fibrillarin and its mobility in living human cells. *J. Cell Biol.* 151:653–662.

Spector, D.L. 2001. Nuclear domains. *J. Cell Sci.* 114:2891–2893.

Staley, J.P., and C. Guthrie. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell.* 92:315–326.

Swedlow, J.R., and A.I. Lamond. 2001. Nuclear dynamics: where genes are and how they got there. *Genome Biol.* 2:0002.

Turley, S.J., E.M. Tan, and K.M. Pollard. 1993. Molecular cloning and sequence analysis of U3 snoRNA-associated mouse fibrillarin. *Biochim. Biophys. Acta.* 1216:119–122.

Verheggen, C., J. Mouaikel, M. Thiry, J.M. Blanchard, D. Tollervey, R. Bordonne, D.L. Lafontaine, and E. Bertrand. 2001. Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. *EMBO J.* 20:5480–5490.

Vidovic, I., S. Nottrott, K. Hartmuth, R. Luhrmann, and R. Ficner. 2000. Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment. *Mol. Cell.* 6:1331–1342.

Visintin, R., and A. Amon. 2000. The nucleolus: the magician's hat for cell cycle tricks. *Curr. Opin. Cell Biol.* 12:372–377.

Warner, J.R. 2001. Nascent ribosomes. *Cell.* 107:133–136.

Watkins, N.J., V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, M. Wilm, M. Rosbash, C. Branlant, and R. Luhrmann. 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell.* 103:457–466.

Weinstein, L.B., and J.A. Steitz. 1999. Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.* 11:378–384.

RNA (2002), *8*:626–636. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838202020022

Yeast Pescadillo is required for multiple activities during 60S ribosomal subunit synthesis

MARLENE OEFFINGER,[1] ANTHONY LEUNG,[2] ANGUS LAMOND,[2] and DAVID TOLLERVEY[1]

[1]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, EH9 3JR, United Kingdom
[2]Wellcome Trust Biocentre, University of Dundee, Dundee, United Kingdom

ABSTRACT

The Pescadillo protein was identified via a developmental defect and implicated in cell cycle progression. Here we report that human Pescadillo and its yeast homolog (Yph1p or Nop7p) are localized to the nucleolus. Depletion of Nop7p leads to nuclear accumulation of pre-60S particles, indicating a defect in subunit export, and it interacts genetically with a tagged form of the ribosomal protein Rpl25p, consistent with a role in subunit assembly. Two pre-rRNA processing pathways generate alternative forms of the 5.8S rRNA, designated $5.8S_L$ and $5.8S_S$. In cells depleted for Nop7p, the $27SA_3$ pre-rRNA accumulated, whereas later processing intermediates and the mature $5.8S_S$ rRNA were depleted. Less depletion was seen for the $5.8S_L$ pathway. TAP-tagged Nop7p coprecipitated precursors to both $5.8S_L$ and $5.8S_S$ but not the mature rRNAs. We conclude that Nop7p is required for efficient exonucleolytic processing of the $27SA_3$ pre-rRNA and has additional functions in 60S subunit assembly and transport. Nop7p is a component of at least three different pre-60S particles, and we propose that it carries out distinct functions in each of these complexes.

Keywords: nucleolus; pre-rRNA; ribosome; RNA processing

INTRODUCTION

Most steps in ribosome synthesis take place within the nucleolus, a specialized subnuclear structure. During ribosome synthesis, a complex processing pathway converts a large pre-rRNA to the mature 18S, 5.8S, and 25S/28S rRNAs (see Fig. 1B). In addition, the mature rRNA sequences within the pre-RNA undergo extensive covalent nucleotide modification and assembly with the 80 ribosomal proteins. More than 80 nonribosomal proteins that are required for ribosome synthesis have been identified by genetic and biochemical approaches in yeast (see Kressler et al., 1999; Venema & Tollervey, 1999; Warner, 2001). Biochemical analyses in human cells have identified an even larger number of nucleolar proteins (Anderson et al., 2002), although in most cases, their function in ribosome synthesis has not yet been directly addressed. Subdomains of the human nucleolus can be identified microscopically. Transcription of the rDNA is believed to occur at the boundaries of the fibrillar centers with initial processing and preribosome assembly occurring in the associated dense fibrillar component (DFC) regions. Later processing and assembly of the pre-ribosomes occurs in the surrounding granular component (GC) of the nucleolus (see, e.g., Shaw & Jordan, 1995; Scheer & Hock, 1999; Lyon & Lamond, 2000). Most analyses of subnuclear structure have been performed on vertebrates and plants, but similar structures are present in yeast (Leger-Silvestre et al., 1997, 1999).

During pre-rRNA processing, the $27SA_2$ pre-rRNA can be processed by two alternative pathways (Henry et al., 1994; see Fig. 1B). In the major pathway, the pre-rRNA is cleaved at site A_3 by RNase MRP, forming the $27SA_3$ pre-rRNA. Subsequent exonuclease digestion to site B_{1S} requires the two known 5′ → 3′ exonucleases, Xrn1p and Rat1p, and generates the 5′ end of the $27SB_S$ pre-rRNA and mature $5.8S_S$ rRNA (Henry et al., 1994). An alternative, poorly understood, pathway processes the pre-rRNA at site B_{1L}, the 5′ end of the $27SB_L$ pre-rRNA. Both 27SB pre-rRNAs are subsequently processed, by apparently identical pathways, to generate the mature 25S rRNA and either the $5.8S_S$ or $5.8S_L$ rRNAs (see Fig. 1). The ratio between the two forms of 5.8S shows some variation between strains, but around 75–80% of the population is normally made up of $5.8S_S$, which is 8 nt shorter than $5.8S_L$. Similar 5′ heterogeneity is seen for 5.8S rRNA from many other Eukaryotes, including humans, *Xenopus*, *Drosophila*, and plants (Henry et al., 1994), sug-

Reprint requests to: David Tollervey, Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, EH9 3JR, United Kingdom; e-mail: d.tollervey@ed.ac.uk.

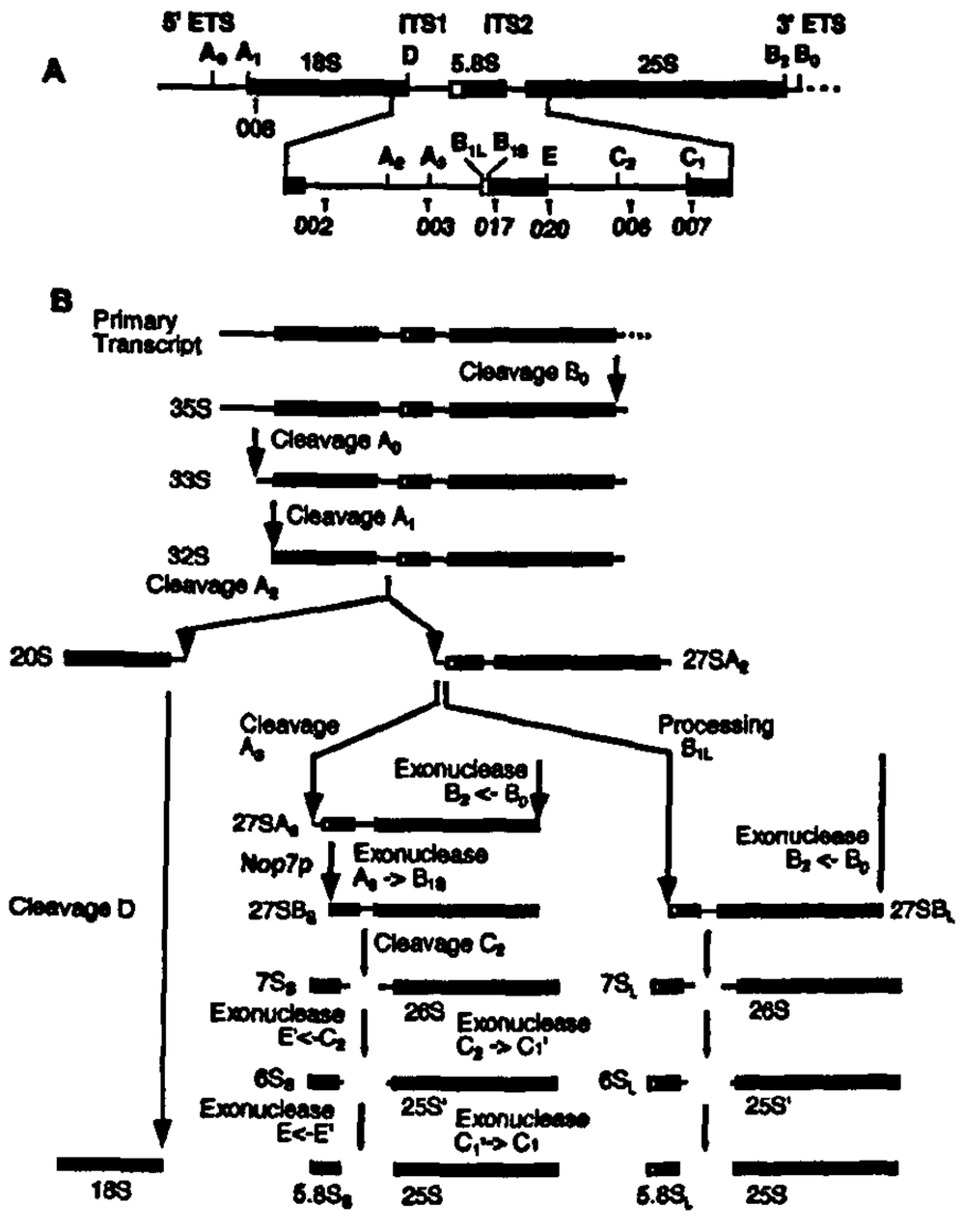


FIGURE 1. Pre-rRNA processing in *S. cerevisiae*. **A:** Structure and processing sites of the 35S pre-rRNA. This precursor contains the sequences for the mature 18S, 5.8S, and 25S, which are separated by the two internal transcribed spacers ITS1 and ITS2 and flanked by the two external transcribed spacers 5'ETS and 3'ETS. The positions of the oligonucleotide probes utilized in northern hybridization and primer extension analyses are indicated. **B:** Pre-rRNA processing pathway. The 35S pre-rRNA is generated by 3′ cleavage at site B_0. 35S is then cleaved at site A_0 to produce the 33S pre-rRNA, which is rapidly cleaved at site A_1, producing the 32S pre-rRNA. 32S is cleaved at site A_2, separating the precursors to the 40S and 60S subunits, the 20S and $27SA_2$ pre-rRNAs, respectively. $27SA_2$ is processed via two alternative pathways. In the major pathway, cleavage at site A_3 by RNase MRP produces $27SA_3$, which is then trimmed to site B_{1S} by the 5′ to 3′ exonucleases Rat1p and Xrn1p, producing the $27SB_S$ pre-rRNA. Alternatively, $27SA_2$ can be processed to $27SB_L$ by an undetermined mechanism. $27SB_S$ and $27SB_L$ are matured to the 5.8S and 25S by identical pathways. Trimming to site B_2 generates the mature 3′ end of the 25S rRNA. Cleavage at site C_2 and exonuclease digestion by Rat1p and Xrn1p generates the 5′ end of mature 25S. The 3′ end of the 5.8S is generated by 3′ to 5′ exonuclease digestion from site C_2 to E. For reviews on pre-rRNA processing and *trans*-acting factors see Kressler et al. (1999), Lafontaine and Tollervey (2001), and Venema and Tollervey (1999).

gesting that the existence of two processing pathways is both conserved and functionally significant.

The Pescadillo gene was initially identified in Zebrafish as the site of retrovirus insertion, which resulted in defects in embryonic development (Allende et al., 1996). Pescadillo mRNA showed widespread expression in developing mouse embryo brain with increased protein levels in replicating cells (Kinoshita et al., 2001). Protein levels were also increased in malignant cells (Kinoshita et al., 2001), possibly related to previous reports of increased nucleolar size and ribosome synthesis in such cells. Pescadillo was localized to the nucleolus in Hela cells (Kinoshita et al., 2001) and the *Schizosaccharomyces pombe* homolog, SPBC19F5.05c, was also found to be nucleolar in a high throughput screen for subcellular localization of GFP fusion proteins (Ding et al., 2000). While this work was in progress, characterization of the yeast Pescadillo homolog Yph1p/Nop7p (YGR103w) was reported. YGR103w was originally published under the name of *YPH1* (Kinoshita et al., 2001), but has been designated as *NOP7* by the *Saccharomyces* genetic database. Nop7p is essential for viability and two temperature-sensitive (ts) lethal mutant alleles were reported to block growth at different steps in the cell-cycle; *yph1-24* led to arrest in G1, whereas the *yph1-45* allele caused G2 arrest (Kinoshita et al., 2001). G1 arrest is expected for mutations defective in ribosome synthesis, which are unable to pass the "Start" checkpoint control, but G2 arrest would not normally be predicted for a ribosome synthesis defect. In addition, Pescadillo was observed to contain a BRCT domain (Haque et al., 2000), which was originally identified in the breast and ovarian cancer gene BRCA1 and has been identified in several proteins involved in cell-cycle checkpoints and DNA repair (reviewed in Bork et al., 1997). Based on these observations, Pescadillo and Yph1p/Nop7p were proposed to perform some cell-cycle specific function.

A proteomic analysis of the human nucleolus identified 271 putative nucleolar proteins including Pescadillo (Anderson et al., 2002), the nucleolar localization of which was confirmed by YFP-tagging. A database search clearly identified YGR103w as the probable yeast homolog and we therefore analyzed its role in ribosome synthesis. While this work was in progress, the purification of a precursor to the 60S ribosomal subunit was reported that made use of a tagged form of Nop7p (Harnpicharnchai et al., 2001). This analysis did not, however, describe the effects of depletion of Yph1p/Nop7p on pre-rRNA processing or ribosome synthesis. Here we show that Nop7p is required for formation of $27SB_S$, and therefore of the mature $5.8S_S$ rRNA, from the $27SA_3$ pre-rRNA and has additional functions in subunit assembly or export.

RESULTS

Human Pescadillo and yeast Nop7p are localized to the nucleolus

A proteomic analysis of purified human nucleoli identified 271 proteins, one of which was Pescadillo (Anderson et al., 2002). To confirm this localization, an eYFP-Pescadillo construct was expressed in Hela cells by transient transfection (Fig. 2A). Comparison of the localization of eYFP-Pescadillo (shown in green; Fig. 2A4) to a DIC image (Fig. 2A1) showed its predominant lo-

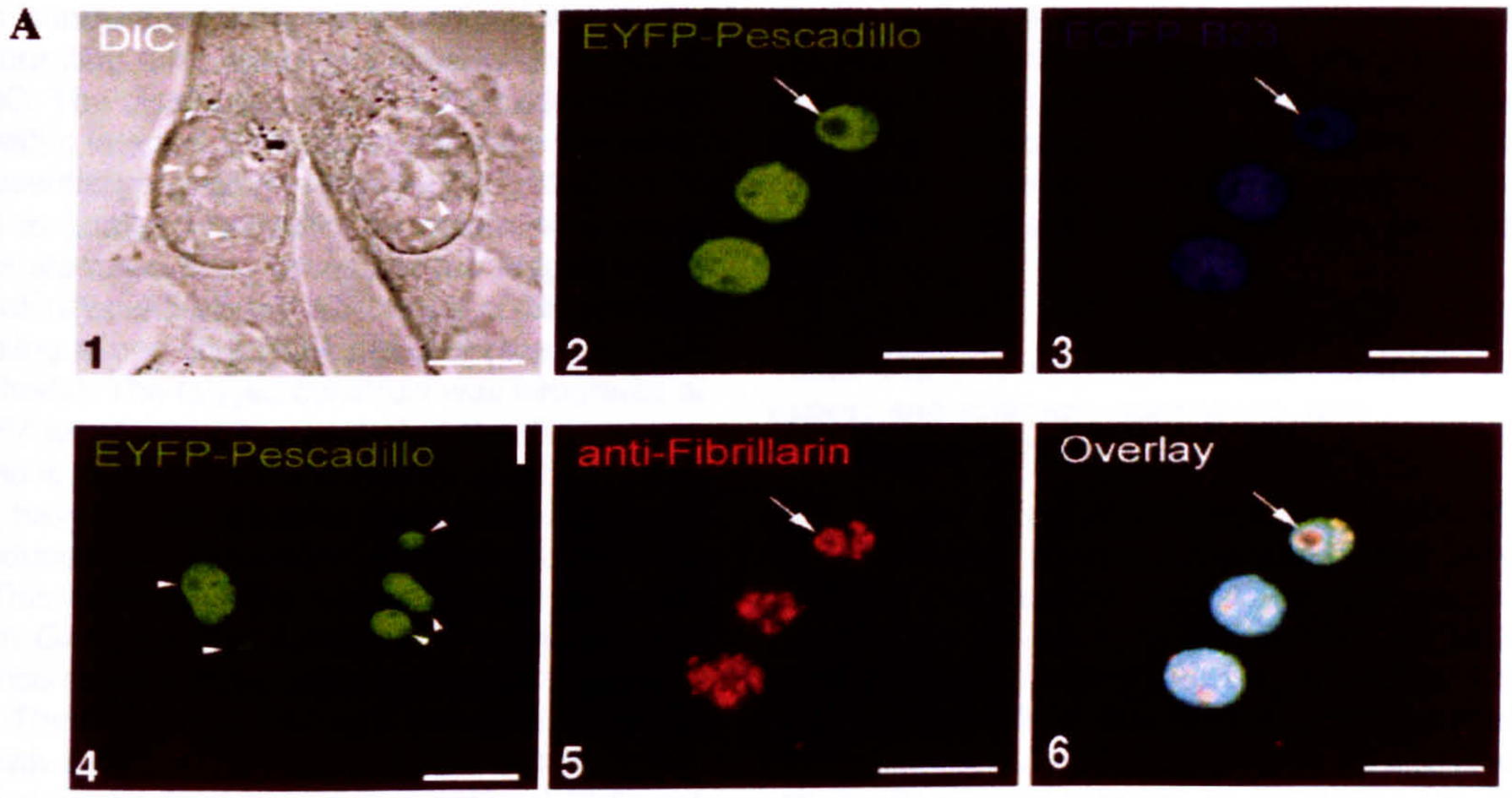


FIGURE 2. Nucleolar localization of Pescadillo and Nop7p. **A:** Localization of human Pescadillo compared with known nucleolar markers. (1, 4) Hela cells were fixed 16 h after transfection with EYFP-Pescadillo. Comparison to the DIC image (1) shows localization of eYFP-Pescadillo (4; green) to nucleoli (arrowheads). 2, 3, 5, and 6: As markers for subnucleolar distribution, the EYFP-Pescadillo transfected cells (2; green) were cotransfected with the granular component protein ECFP-B23 (3; blue) and decorated with antibodies directed against the dense fibrillar component (DFC) protein fibrillarin (5; red). Scale bar = 5 μm. **B:** Localization of Nop7p. The *GAL::nop7-TAP* strain also expressing the nucleolar marker DsRedNop1p was examined by indirect immunofluorescence using an anti-protein A antibody coupled to FITC. Also shown is the position of the nucleus visualized by DAPI staining and a wild-type control strain.

calization to nucleoli (indicated by arrowheads) with a low level of nucleoplasmic staining.

The subnucleolar distribution of eYFP-Pescadillo (Fig. 2A2) was compared to the GC marker eCFP-tagged B23/nucleophosmin (Npm1) (shown in blue; Fig. 2A3) and the DFC marker fibrillarin (shown in red; Fig. 2A5). Fibrillarin is a component of the box C+D snoRNAs (Schimmang et al., 1989) that function early in ribosome synthesis, whereas B23 is a putative assembly factor and nuclease that is believed to act later in ribosome synthesis (Biggiogera et al., 1990; Savkur & Olson, 1998). In vitro, B23 is reported to cleave a pre-rRNA reporter within ITS2 (Savkur & Olson, 1998), at a site potentially equivalent to C_2 in the yeast pre-rRNA. Fibrillarin is concentrated in the DFC whereas B23 was reported to localize to the periphery of the DFC and the GC based on immuno-EM (Biggiogera et al., 1990), consistent with a later role for B23 in nucleolar ribosome maturation. The distribution of eYFP-Pescadillo resembled that of eCFP-B23, but was distinct from that of fibrillarin. eYFP-Pescadillo and eCFP-B23 were largely excluded from the DFCs (one

of which is indicated by an arrow) and concentrated in the surrounding area, which presumably corresponds to the GC. The distribution of eYFP-Pescadillo is consistent with a late role in nucleolar ribosome synthesis.

The essential yeast protein Nop7p (YGR103w) is 40% identical to human Pescadillo. To determine whether Nop7p is also nucleolar, it was epitope tagged with a tandem-affinity purification (TAP) construct (Rigaut et al., 1999) using a one-step PCR protocol (see Materials and Methods). The tagged construct was integrated at the *NOP7* locus under the control of the *GAL10* promoter and is the only source of Nop7p. The host strain, YDL401, has reduced galactose permease activity leading to reduced *GAL* induction (Lafontaine & Tollervey, 1996). This eliminates the overexpression generally seen with *GAL*-regulated constructs and allows faster appearance of phenotypes following transfer to glucose medium. The *GAL::nop7-TAP* cells exhibited no detectable growth defect on permissive RSG medium, showing the fusion construct to be fully functional (data not shown).

To determine the location of Nop7-TAP, cells were examined by indirect immunofluorescence (Fig. 2B) using a rabbit anti-protein A and a secondary FITC-coupled goat anti-rabbit antibody to detect the protein A region of the TAP tag (Rigaut et al., 1999). As a marker for the nucleolus, a DsRed fusion with the nucleolar protein Nop1p (the yeast homolog of fibrillarin) was coexpressed as previously described (Gadal et al., 2001b), and the nucleoplasm was identified by DAPI staining. Anti-protein A preferentially decorated the nucleolus, with a weaker signal over the nucleoplasm. No cytoplasmic signal was detected. We conclude that Nop7p is localized to the nucleus with nucleolar enrichment. The significant nucleoplasmic staining would be consistent with association with late pre-ribosomes that have been released from the nucleolus (see Milkereit et al., 2001).

Yeast Nop7p is required for 60S subunit export and interacts genetically with GFP-tagged Rpl25p

To examine the possible functions of Nop7p in ribosome synthesis, its expression was placed under the control of a repressible *GAL10* promoter using a one-step PCR technique in strain YDL401 (see Materials and Methods). Growth of the *GAL::nop7* strain was not clearly different from the isogenic wild-type strain on RGS medium, but was progressively slowed following transfer to glucose medium, commencing around 6 h after transfer (Fig. 3A). The *yph1-45* allele of *NOP7* is reported to lead to a G2 arrest phenotype, consistent with a specific cell-cycle defect (Kinoshita et al., 2001). However, microscopic inspection of the *GAL::nop7* strain following transfer to glucose medium showed only the accumulation of unbudded cells, even after 24 h, indicating arrest in G1 (data not shown). This is the expected phenotype for a defect in ribosome synthesis that results in the inability to pass the "Start" checkpoint.

Several recent studies have made use of fusions between ribosomal proteins and GFP to follow the ex-

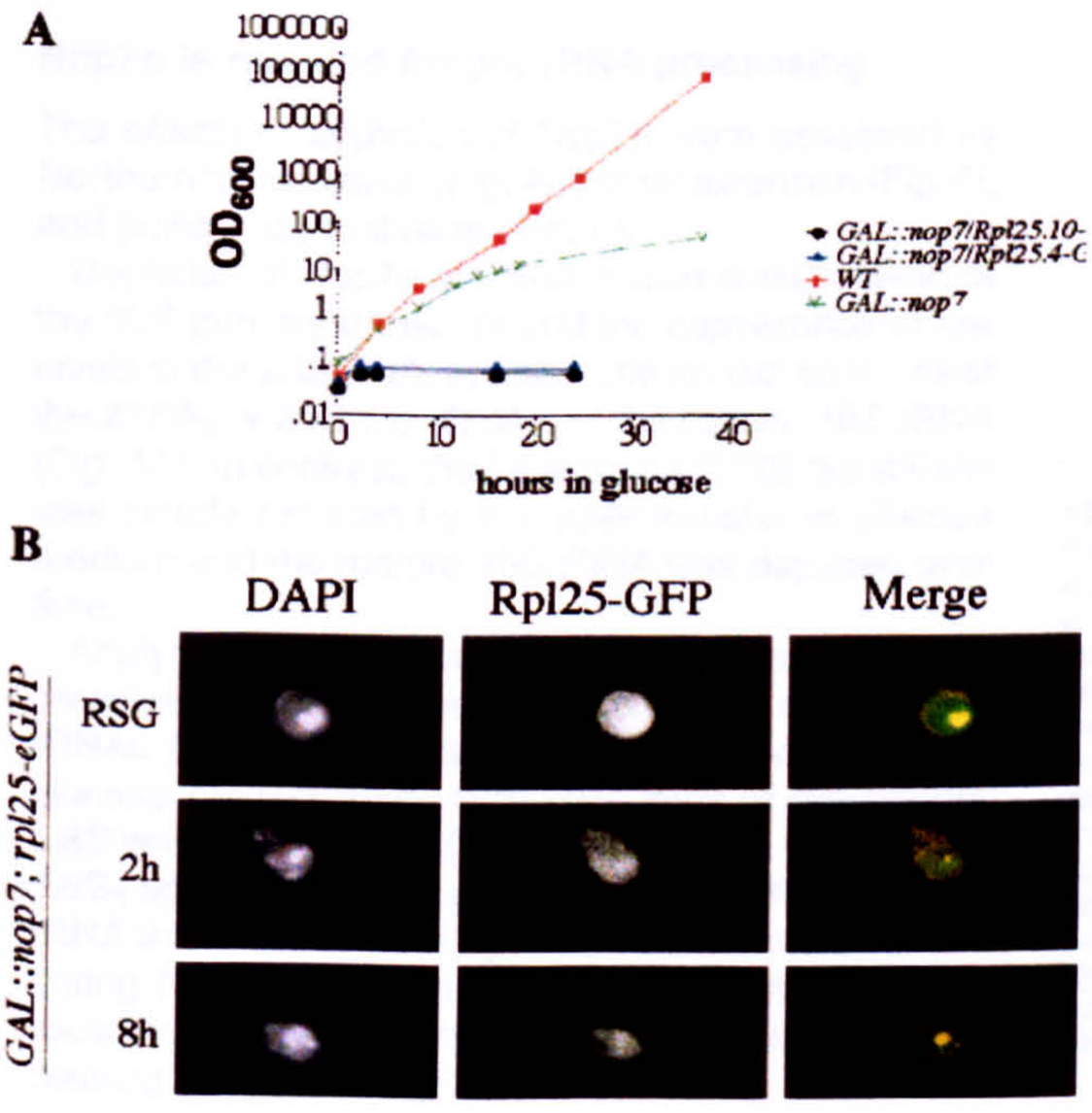


FIGURE 3. Nop7p is required for 60S subunit export and interacts genetically with GFP-tagged Rpl25p. **A:** Growth curves of *GAL::nop7* strains following transfer to glucose medium, with and without expression of Rpl25p-eGFP. Strains were pregrown in RGS medium and transferred to glucose medium for the times indicated. Strains were maintained in exponential growth by dilution with prewarmed medium. Cell densities measured by OD_{600} are shown corrected for dilution. (■) Wild-type; (∗) *GAL::nop7*; (●) *GAL::nop7*; *rpl25.10-GFP*; (▲) *GAL::nop7*; *rpl25.4-GFP*. **B:** Subcellular distribution of Rpl25-eGFP in a *GAL::nop7* strain. Rpl25-eGFP was examined by fluorescence microscopy during growth in RGS medium and following transfer to glucose medium for 2 and 8 h. The position of the nucleus was visualized by DAPI staining, which also stains the cytoplasm more weakly due to the presence of mitochondria. In the merged image, DAPI staining is shown in red and Rpl25-GFP is in green.

port of 60S ribosomal subunits from the nucleus to the cytoplasm (Stage-Zimmermann et al., 2000; Baßler et al., 2001; Gadal et al., 2001a, 2001b; Milkereit et al., 2001; Fatica et al., 2002). To look for 60S subunit export defects, Rpl25p-eGFP (Gadal et al., 2001b) was expressed from a plasmid in the wild-type and *GAL::nop7* strains. As previously reported, expression of this construct had little effect on the growth of the wild-type strain (Gadal et al., 2001b) or the *GAL::nop7* strain on galactose medium (data not shown). Unexpectedly, growth of the *GAL::nop7*/Rpl25p-eGFP strain was very rapidly inhibited following transfer to glucose medium (Fig. 3; two independent transformants are shown). These strains also express the wild-type Rpl25p, showing that the Rpl25p-eGFP fusion is dominant negative for growth in strains with a reduced level of Nop7p. The growth inhibition is much more rapid than would have been expected for a strain that is simply unable to synthesize new ribosomes (see Discussion) and we conclude that Nop7p has a role in 60S ribosomal subunit assembly.

The distribution of Rpl25p-eGFP was followed during depletion of Nop7p (Fig. 3B). During growth of the *GAL::nop7* strain on RSG medium, Rpl25-eGFP showed the normal, predominantly cytoplasmic distribution. After transfer to glucose medium for 2 h, increased nuclear staining of Rpl25-eGFP was already visible, and accumulation was strong after 8 h. The distribution of Rpl25-eGFP fluorescence matched that of DAPI staining, indicating that it was not restricted to the nucleolus. We conclude that Nop7p is required to allow the export of precursors to the 60S ribosomal subunit from the nucleoplasm to the cytoplasm.

Nop7p is required for pre-rRNA processing

The effects of depletion of Nop7p were assessed by Northern hybridization (Fig. 4), primer extension (Fig. 5), and pulse-chase labeling (Fig. 6).

Depletion of Nop7p resulted in mild accumulation of the 35S primary transcript and the appearance of low levels of the 23S RNA, but had little impact on levels of the $27SA_2$ or 20S pre-rRNAs, or the mature 18S rRNA (Fig. 4A). In contrast, the level of the 27SB pre-rRNAs was clearly reduced by 8 h after transfer to glucose medium and the mature 25S rRNA was depleted over time.

Analysis of low-molecular-weight RNAs showed progressive reduction in the levels of the 7S and 6S pre-rRNAs following transfer of the *GAL::nop7* strain to glucose medium (Fig. 4Bb). The level of the mature 5.8S was also reduced (Fig. 4Bc), and the reduction in $5.8S_S$ appeared slightly greater than for $5.8S_L$. The pre-rRNA that extends from A_2 to C_2 was not accumulated during Nop7p depletion (Fig. 4Ba), in contrast to the recently reported effects of depletion of another processing factor, Ssf1p (Fatica et al., 2002).

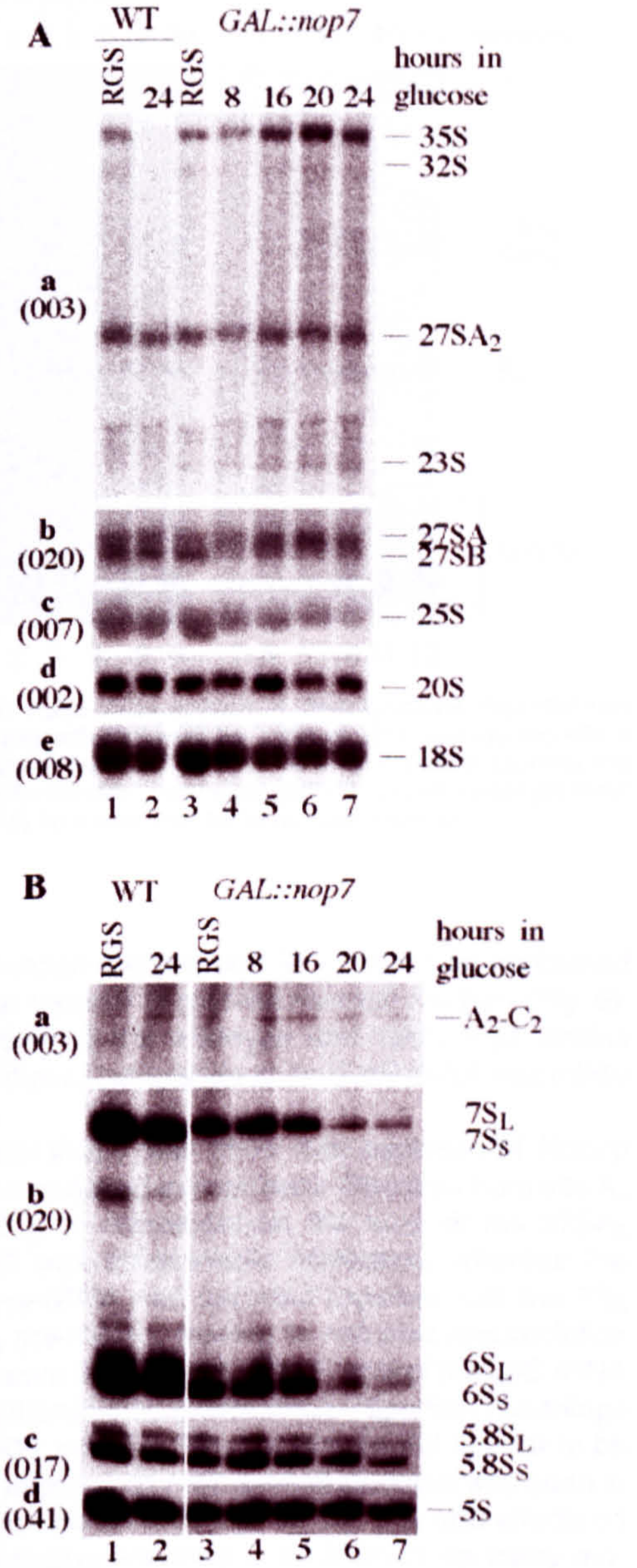


FIGURE 4. Northern analysis of the effects of Nop7p depletion on pre-rRNA processing. Lanes 1 and 2: wild-type strain in RGS medium and 24 h after transfer to glucose. Lanes 3–7: *GAL::nop7* strain in RGS medium and after transfer to glucose medium for the times indicated. **A:** (a) Hybridization of probe 003, complementary to ITS1 upstream of A_3. (b) Hybridization with probe 020, complementary to the 5.8S/ITS2 boundary. (c) Hybridization with probe 007, complementary to the 25S rRNA. (d) Hybridization with probe 002, complementary to ITS1 upstream of A_2. (e) Hybridization with probe 008, complementary to 18S rRNA. **B:** (a) Hybridization with probe 003, complementary to ITS1 upstream of A_3. (b) Hybridization with probe 020, complementary to the 5.8S/ITS2 boundary. (c) Hybridization with probe 017, complementary to 5.8S rRNA. (d) Hybridization with probe 041, complementary to 5S rRNA. RNA was separated on a 1.2% agarose/formaldehyde gel (**A**) or 8% polyacrylamide/urea gel (**B**). Probe names are indicated in parentheses on the left.

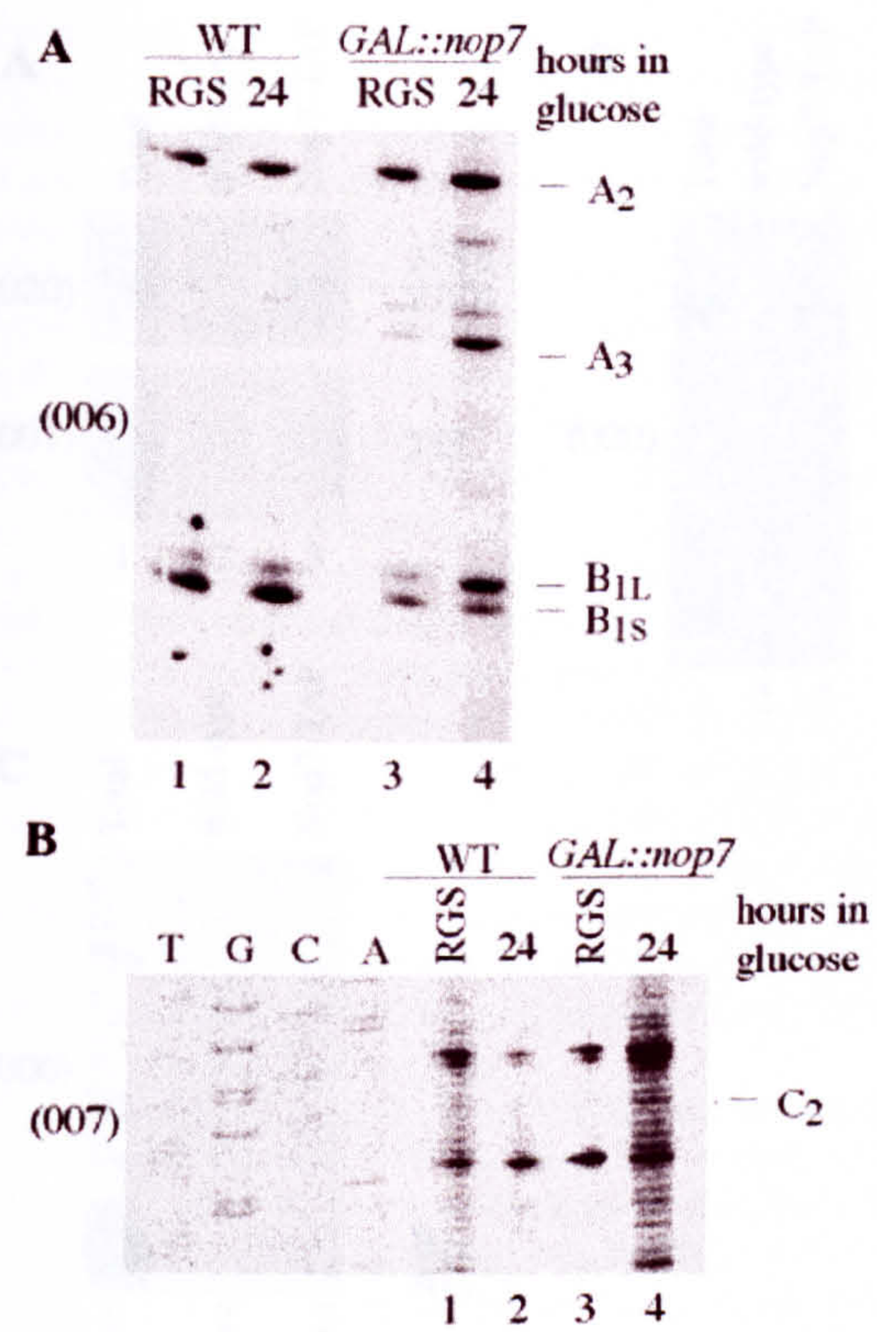


FIGURE 5. Primer extension analysis of pre-rRNA processing. Lanes 1 and 2: wild-type strain in RGS medium and 24 h after transfer to glucose medium. Lanes 3 and 4: *GAL::nop7* strain in RGS medium and 24 h after transfer to glucose medium. **A:** Primer extension using oligo 006, which hybridizes within ITS2, 3′ to site C_2. Primer extension stops at sites A_2, A_3, B_{1S}, and B_{1L} show the levels of the $27SA_2$, $27SA_3$, $27SB_L$, and $27SB_S$ pre-rRNAs, respectively. **B:** Primer extension using oligo 007, which hybridizes within 25S rRNA. The primer extension stop at C_2 shows the level of the 26S pre-rRNA.

$5.8S_S$ is processed from the $27SB_S$ pre-rRNA, whereas $5.8S_L$ is processed from $27SB_L$ (see Fig. 1B). To determine the levels of the 27SB species, they were analyzed by primer extension (Fig. 5A) using an oligo hybridizing within the 3′ region of ITS2 (oligo 006; see Fig. 1A). Following growth of the *GAL::nop7* strain on glucose medium, the level of $27SB_S$ was clearly reduced relative to $27SB_L$, as shown by the primer extension stops at B_{1S} and B_{1L}, respectively (Fig. 5A, lane 4). Consistent with the northern analysis, little alteration was seen in the level of $27SA_2$, as shown by the primer extension stop at site A_2. In contrast, the level of $27SA_3$, shown by the stop at site A_3, was substantially elevated. Using a primer hybridizing within the mature 25S rRNA (oligo 007; see Fig. 1A), slight accumulation was seen for the primer extension stop at site C_2, the 5′ end of the 26S pre-rRNA (Fig. 5B). This effect was weak, however, and its significance is unclear.

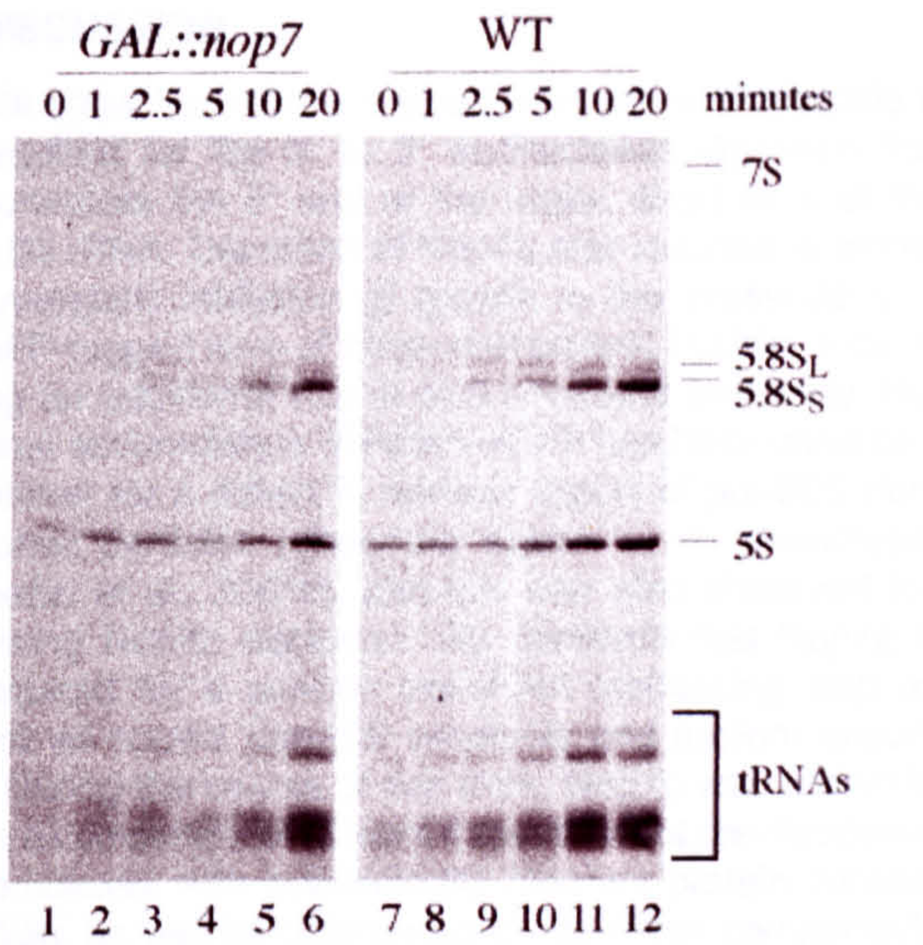


FIGURE 6. Pulse-chase analysis of rRNA synthesis. Pre-rRNA was pulse labeled with [^{3}H]uracil for 2 min at 30 °C and chased with a large excess of unlabeled uracil for the times indicated. Labeling was performed for the *GAL::nop7* strain (lanes 1–6) and a wild-type strain (lanes 7–12) 16 h after transfer to glucose medium.

Pulse-chase analysis with [H^3]-uracil was performed 16 h after transfer to glucose minimal medium (Fig. 6). Comparison of the wild-type and *GAL::nop7* strains showed that accumulation of the 5.8S rRNA was mildly delayed.

Together these data show that depletion of Nop7p resulted in reduced exonuclease digestion from site A_3 to site B_{1S}. In consequence, the level of the $27SA_3$ pre-rRNA was substantially increased, whereas the $27SB_S$ pre-rRNA was depleted together with the $7S_S$ and $6S_S$ pre-rRNAs, leading to reduced accumulation of the mature $5.8S_S$ rRNA. The 5′ end of the 25S rRNA is also generated by exonuclease digestion (Geerlings et al., 2000; see Fig. 1B), but this did not appear to be strongly affected, as only a small increase was seen in the primer extension stop at site C_2. The mild effects on 35S processing are likely to be indirect, as many mutations that inhibit synthesis of 60S subunits result in partial inhibition of the early pre-rRNA processing steps (for further discussion see Venema & Tollervey, 1999).

Nop7p is associated with pre-rRNAs from both processing pathways

To determine whether Nop7p associated specifically with the $27SB_S$ branch of the processing pathway, coprecipitated RNAs were analyzed by northern analysis and primer extension. Northern hybridization (Fig. 7A,B) showed that the 27SB and 7S pre-rRNAs coprecipitated with Nop7-TAP, but were not detectably recov-

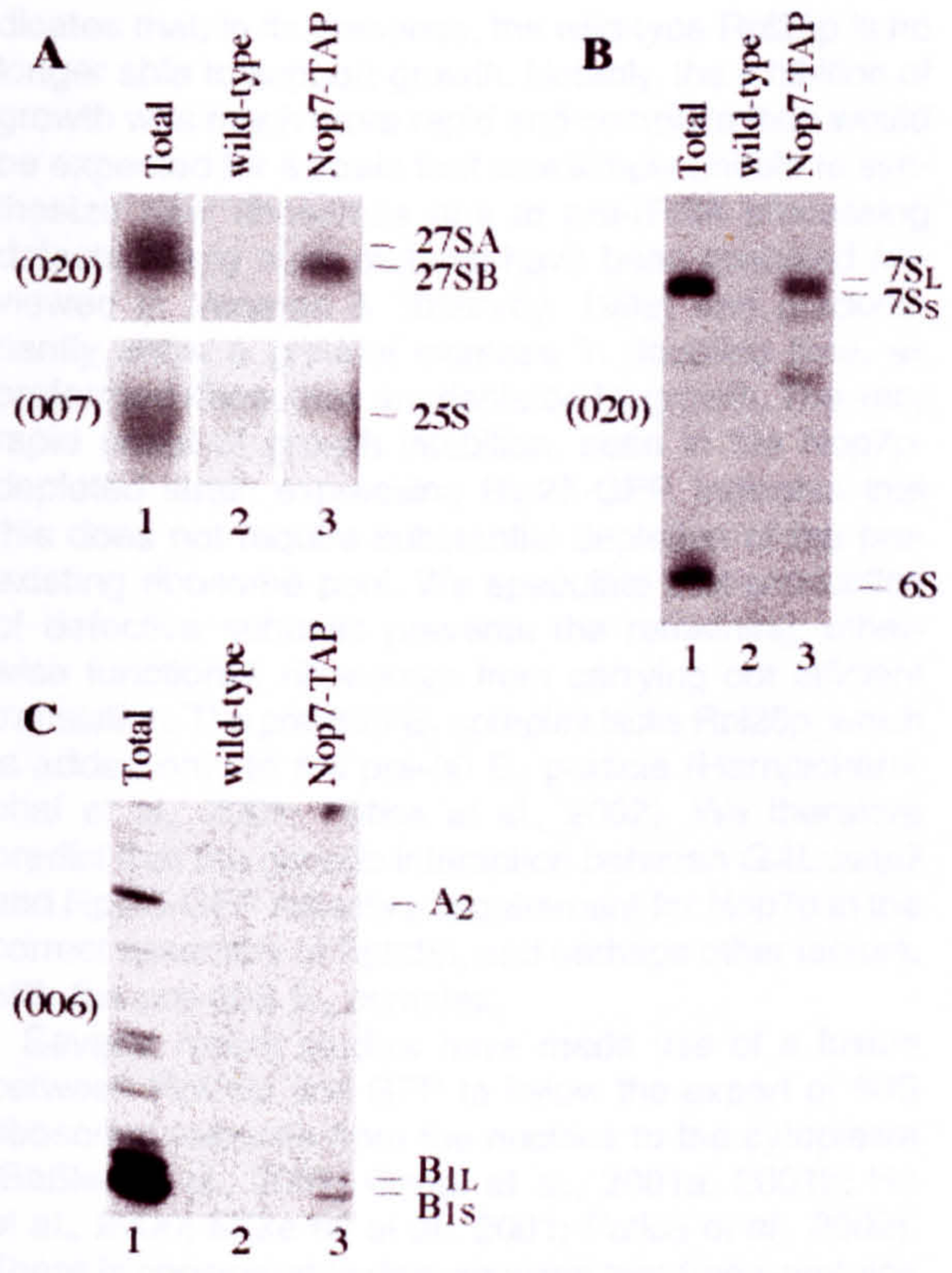


FIGURE 7. Analysis of RNAs coprecipitated with TAP-tagged Nop7p. Lane 1: Total RNA control (5 μg). Lane 2: Precipitate from a wild-type control strain. Lane 3: Precipitate from a strain expressing Nop7-TAP. **A:** Northern hybridization of high-molecular-weight RNA separated on a 1.2% agarose/formaldehyde gel. **B:** Northern hybridization of low-molecular-weight RNA separated on an 8% polyacrylamide/urea gel. **C:** Primer extension analysis. Nop7-TAP was immunoprecipitated from cell lysates using IgG agarose, with release of bound RNA–protein complexes by cleavage of the protein A linker by TEV protease. RNA was recovered from the released material, and from a mock-treated, isogenic wild-type control strain. Oligonucleotides used are indicated in parentheses. The preparation used in **C** is different from that used for **A** and gave lower recovery efficiency.

ered in the mock precipitation from the nontagged wild-type strain. In contrast, the $27SA_2$ and 6S pre-rRNAs were not detectably coprecipitated. The 25S rRNA gave the same background signal in both the wild-type and Nop7-TAP precipitates. Inspection of the original figure showed that both the $7S_L$ and $7S_S$ pre-rRNAs were coprecipitated. $27SB_L$ and $27SB_S$ cannot be resolved by northern hybridization, but primer extension from oligo 006 in ITS2 (see Fig. 1A) showed that both the $27SB_L$ and $27SB_S$ pre-RNAs were coprecipitated with Nop7-TAP (Fig. 7C).

Nop7p has a specific role in formation of the $27SB_S$ pre-rRNA but is associated with pre-rRNAs in both processing pathways, consistent with the conclusion that it has additional roles in 60S subunit assembly and export.

DISCUSSION

We show here that the yeast homolog of Pescadillo is required for the 5′ to 3′ exonuclease digestion that generates the 5′ end of the major, short form of the 5.8S rRNA. Depletion of Nop7p also resulted in strong synergistic inhibition of growth in the presence of a GFP-tagged form of ribosomal protein Rpl25p, indicating an additional role in 60S ribosome assembly. Nuclear accumulation of Rpl25-eGFP has been used as a marker for a defect in nuclear export of pre-60S ribosomal particles (ribosome export or *rix* phenotype; Gadal et al., 2001b), and this was also observed following Nop7p depletion. We conclude that Nop7p is required for a specific pre-rRNA processing step as well as correct pre-60S assembly and nuclear export.

During the course of this work, Nop7p was shown to be a component of at least three different pre-ribosomal complexes with substantially different protein composition, as well as differences in pre-rRNA components (Baßler et al., 2001; Harnpicharnchai et al., 2001; Fatica et al. 2002; see Fig. 8). These analyses allow us to propose a correlation between the pre-ribosomal particles with which Nop7p is associated and the distinct defects in ribosome synthesis that are seen on its depletion.

The earliest pre-60S particle with which Nop7p is known to be associated is pre-60S E_1. This complex is also associated with the $27SA_2$, $27SA_3$, and 27SB pre-rRNAs (Fatica et al., 2002) and it is therefore very likely that Nop7p is required for processing from $27SA_3$ to 27SB within the pre-60S E_1 particle.

A fast acting, dominant negative phenotype is associated with the expression of a GFP-tagged form of the ribosomal protein Rpl25p in strains depleted of Nop7p. The fact that expression of Rpl25-GFP is dominant in-

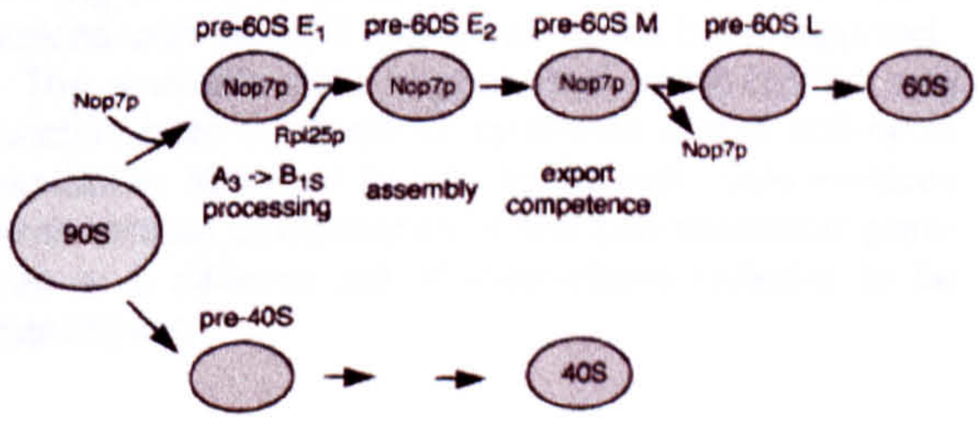


FIGURE 8. Model for the roles of Nop7p in 60S subunit biogenesis. Outline pathway of biogenesis of 60S and 40S ribosomal subunits, modified from Fatica et al. (2002). This model indicates the presence of Nop7p in three different pre-60S complexes designated E_1, E_2, and M, which can be correlated with the different functions deduced for Nop7p. Pre-60S E_1 contains the $27SA_3$ pre-rRNA, the processing of which is defective in strains lacking Nop7p. Rpl25p is not present in pre-60S E_1, but joins the pre-60S E_2 particle, and the defect in Rpl25p assembly is therefore predicted to occur at this step. The pre-60S M complex contains numerous factors required for 60S subunit export as judged by the nuclear retention of a Rpl25-GFP reporter construct, and Nop7p is likely to be required during the acquisition of export competence within this complex.

dicates that, in its presence, the wild-type Rpl25p is no longer able to support growth. Notably, the inhibition of growth was much more rapid and complete than would be expected for a strain that was simply unable to synthesize new ribosomes due to pre-rRNA processing defects. Many such mutants have been analyzed (reviewed in Venema & Tollervey, 1999) and predominantly show a gradual increase in doubling time, as preformed ribosomes are depleted by growth. The very rapid onset of growth inhibition, seen in the Nop7p-depleted strain expressing Rlp25-GFP, indicates that this does not require substantial depletion of the pre-existing ribosome pool. We speculate that production of defective subunits prevents the remaining, otherwise functional, ribosomes from carrying out efficient translation. The pre-60S E_1 complex lacks Rpl25p, which is added only to the pre-60 E_2 particle (Harnpicharnchai et al., 2001; Fatica et al., 2002). We therefore predict that the genetic interaction between *GAL::nop7* and Rpl25-GFP reflects a requirement for Nop7p in the correct assembly of Rpl25p, and perhaps other factors, with the pre-60S E_2 complex.

Several recent studies have made use of a fusion between Rpl25p and GFP to follow the export of 60S ribosomal subunits from the nucleus to the cytoplasm (Baßler et al., 2001; Gadal et al., 2001a, 2001b; Ho et al., 2000; Milkereit et al., 2001; Fatica et al., 2002). There is considerable data showing that free r-proteins do not accumulate in the absence of ribosome synthesis. The accumulation of Rlp25-GFP has therefore been taken as evidence for the accumulation of pre-ribosomal particles in the nucleoplasm, indicating a defect in their export. This assay has defined a late pre-ribosomal particle (pre-60S M in Fig. 8), all tested components of which are required for 60S subunit export. These include Nug1p, Nug2p, Noc2p, Noc3p, and Rix1p as well as Nop7p itself (Baßler et al., 2001; Gadal et al., 2001a, 2001b; Milkereit et al., 2001). Mutations in any of these proteins leads to defects in export, suggesting a requirement for the intact structure of this pre-ribosomal particle. Because multiple components of this complex are required for subunit export, we predict that export competence is established within this particle, and that this activity requires Nop7p.

Mutations in Nug1p, Nug2p, Noc2p, Noc3p, or Rix1p did not result in pre-rRNA processing defects similar to Nop7p depletion or synergistic interactions with Rpl25-GFP (Baßler et al., 2001; Milkereit et al., 2001) indicating that these are distinct activities. Moreover, Nug1-TAP did not coprecipitate $27SA_2$ or $27SA_3$ (Baßler et al., 2001; E. Petfalski & D. Tollervey, unpubl. observations) indicating that it associates with the pre-rRNA particle only after processing at these sites is complete. Depletion of a specific component of the pre-60S E_1 complex, Ssf1p, also did not interact genetically with Rpl25-GFP and did not inhibit subunit export as judged by nuclear accumulation of Rpl25-GFP (Fatica et al., 2002). We therefore propose that the roles of Nop7p in pre-rRNA processing, assembly, and export are distinct and performed within different pre-ribosomal particles (see Fig. 8).

Pescadillo is a multifunctional protein

Pescadillo was isolated as a mutation affecting embryonic development (Allende et al., 1996) and a mutant allele of the yeast gene resulted in growth arrest in G2 (Kinoshita et al., 2001), consistent with a specific defect in cell-cycle progression. Yeast Yhp1p/Nop7p is also reported to interact with Yvh1p (Sakumoto et al., 2001), a protein-tyrosine phosphatase with a postulated role in the regulation of sporulation and meiosis.

There are clear precedents for proteins that function both in cell-cycle progression and ribosome synthesis. Exit from mitosis in budding yeast requires a group of proteins, including the phosphatase Cdc14p, which down-regulate cyclin-dependent kinase activity. Cdc14p is sequestered in the nucleolus in the RENT (regulator of nucleolar silencing and telophase) complex with Sir2p and Net1p, which serves to anchor the complex (Shou et al., 1999). In addition, Net1p is required for the maintenance of normal nucleolar structure and its binding stimulates RNA polymerase I (Shou et al., 1999, 2001). These nucleolus-specific functions of Net1p can be separated genetically from its cell-cycle functions in the RENT complex. In human cells, the nucleolar p14/ARF protein binds and sequesters the negative regulator of p53 activity, Mdm2 (Tao & Levine, 1999; Weber et al., 1999; Zhang & Xiong, 1999). Free Mdm2 ubiquitinates p53 and transports it to the cytoplasm where it is degraded by the proteosome (Fuchs et al., 1998; Geyer et al., 2000), and the nucleolar sequestration of Mdm2 contributes to the inhibition of this activity by ARF. Mouse Pescadillo was identified by its up-regulation in cells lacking p53 (Kinoshita et al., 2001), but other interactions with the p53 system have not been reported.

The available data suggest that yeast Nop7p may function both in ribosome synthesis and in cell-cycle regulation. Whether its role in the cell cycle involves other protein components of the pre-ribosomal particles or a different set of interactions remains to be determined.

MATERIALS AND METHODS

Strains

Growth and handling of *Saccharomyces cerevisiae* were by standard techniques. *GAL*-regulated strains were pregrown in RGS medium, containing 2% raffinose, 2% galactose, and 2% sucrose, and harvested at intervals following a shift to medium containing 2% glucose. Strains for pulse-chase analysis were pregrown in minimal RGS medium lacking uracil and shifted to minimal glucose medium lacking uracil. Strains for

immunofluorescence studies were grown in minimal glucose medium lacking leucine.

Yeast strains used and constructed in this study are listed in Table 1. Conditional mutants under the control of the repressible *GAL10* promoter were generated by one-step PCR strategy in the strains YDL401 and BMA64 (Lafontaine & Tollervey, 1996). Transformants were selected for HIS+ prototrophy and screened by PCR. TAP-tagged strains were constructed by one-step PCR strategy in the GAL-mediated strain *GAL::nop7* (Rigaut et al., 1999). Transformants were screened by immunoblotting and PCR.

TAP-tagged strains were transformed with pUN100Ds RedNOP1 (kindly provided by E. Hurt and U. Heidelberg) to allow ready identification of the nucleolus, and pYEplac195-L25-eGFP to look at nuclear export of 60S ribosomal subunits.

For construction of eYFP-PES, complementary DNA of human Pescadillo gene (GI:2194202) was isolated by PCR amplification from Marathon-Ready Hela cDNA library (Clontech) using specific primers with *Bgl*II and *Eco*RI restriction sites attached to the 5′ and 3′ primer, respectively. The amplified fragment was subsequently cloned to the *Bgl*II-*Eco*RI fragment of eYFP-C1 and verified by DNA sequencing.

RNA extraction, northern hybridization, and primer extension

RNA was extracted as described previously (Tollervey & Mattaj, 1987). For high-molecular-weight RNA analysis, 7 μg of total RNA were separated on a 1.2% agarose gel containing formaldehyde and transferred for northern hybridization as described previously (Tollervey, 1987). Standard 6% or 8% acrylamide-8 M urea gels were used to analyze low-molecular-weight RNA species and primer extension reactions. Primer extensions were performed as described previously (Beltrame & Tollervey, 1992) on 5 μg of total RNA using primers 007 and 006.

For pre-rRNA and rRNA analysis the following oligonucleotides were used:

002: 5′-GCTCTTTGCTCTTGCC;
003: 5′-TGTTACCTCTGGGCCC;
006: 5′-AGATTAGCCGCAGTTGG
007: 5′-CTCCGCTTATTGATATGC;
008: 5′-CATGGCTTAATCTTTGAGAC;
017: 5′-GCGTTGTTCATCGATGC;
020: 5′-TGAGAAGGAAATGACGCT;
041: 5′-CTACTCGGTCAGGCTC.

Immunofluorescence

For localization of yeast Nop7p, cells were grown in selective medium to midexponential phase, fixed by incubation in 4% (v/v) formaldehyde for 30 min at 25 °C, and spheroplasted. Immunofluorescence was then performed as described previously (Grandi et al., 1993; Bergès et al., 1994). Protein A fusions were detected with a rabbit anti-Protein A antibody (Sigma) and a secondary goat anti-rabbit antibody coupled to FITC (Sigma) at 1:1,000 and 1:200 dilutions, respectively. To stain nuclear DNA, DAPI was included in the mounting medium (Vectashield, Vector Laboratories). Cells were viewed on a Zeiss Axioscop microscope.

Cells containing pYE195-Rpl25-eGFP were grown in SD-LEU to midexponential phase, fixed in 4% (v/v) formaldehyde for 30 min, and pelleted. Cells were resuspended in 100 mM KH_2Ac/K_2HAc/1.1 M sorbitol and mounted onto slides using moviol, containing DAPI. To detect Rpl25-eGFP in vivo in the fluorescence microscope, the GFP-signal was examined in the fluorescein channel of a Zeiss Axioscop microscope (Hurt et al., 1999). Pictures were obtained with SmartCapture VP.

The localization of eYFP-Pescadillo was determined after transient transfection into Hela cells. EYFP-PES and eCFP-B23 were cotransfected for 6 h using Effectene (Quiagen) according to the manufacturer's protocol and fixed after 42 h using 3.7% paraformaldehyde in CSK buffer. Cells were permeabilized and decorated with antibodies against dense fibrillar component marker fibrillarin (72B9) and the granular component marker B23 (anti-B23). Cells were imaged using a Zeiss LSM410 confocal microscope or a Zeiss DeltaVision Restoration microscope (Applied Precision, Inc.). Images presented here are maximal projections of the entire nuclear fluorescence.

Immunoprecipitation of *GAL::nop7*-TAP

For immunoprecipitation of *GAL::nop7*-TAP, cells were grown in YPgal to OD_{600} = 2 and lysed in buffer A (150 mM KAc, 20 mM Tris-Ac, pH 7.5, 5 mM MgAc) with 1 mM DTT, 0.5% Triton X-100, 2.5 mM vanadyl-ribonucleoside complexes (VRC), and 5 mM PMSF (phenylmethylsulphonylfluoride) at

TABLE 1. Yeast strains used and constructed in this study.

Strain	Genotype	Reference
YDL401	MATa *his3Δ200 leu2Δ1 trp1 ura3-52 gal2 galΔ108*	Lafontaine & Tollervey, 1996
YMO1	as YDL401 but *GAL10::nop7-HIS3*	This work
YMO2	as YDL401 but *GAL10::nop7-TAP-TRP1*	This work
YMO3	as YMO2 but pUN100-DsRed*nop1* LEU1	This work
YMO4	as YMO1 but pRS315-*Rpl25*-eGFP	This work
BMA64	MATa *ade2-1 his3-11,-15 leu2-3,-112 trp1Δ, ura3-1*	F. Lacroute
YMO5	as BMA64 but pA3*ura*3	This work
YMO6	as BMA64 but *GAL10::nop7*, pA3*ura*3	This work
YCA31	as YDL401 but *GAL10::prot.A-RRP4, RRP6:(Kl)TRP1*	Allmang et al., 1999
GAL::DOB1	MATa *ura3-1 ade2-1 his3-11,-15 leu2-3,-112 trp1-1* Dob1::HIS3MX6 +[pAS24-*DOB1*]	de la Cruz et al., 1998

4 °C using glass beads (Sigma). Immunoprecipitation with rabbit IgG agarose beads and subsequent TEV cleavage were performed as described (Rigaut et al., 1999). RNA was extracted with buffer AE/phenol-chloroform, ethanol precipitated (Schmitt et al., 1990), and analyzed by northern hybridization and primer extension.

Pulse-chase labeling experiments

Pulse-chase labeling of pre-rRNA was performed as previously described (Tollervey et al., 1993) using 100 μCi [5,6-^{3}H]uracil (Amersham) for 2 min at 30 °C. Unlabeled uracil was added to a final concentration of 240 μgmL^{-1}. Samples (1 mL) were taken, transferred to microcentrifuge tubes at room temperature, and spun for 10 s at full speed in an Eppendorf centrifuge. Cell pellets were frozen in liquid N_2. Total RNA was extracted with buffer AE/phenol-chloroform and ethanol precipitated (Schmitt et al., 1990). [H^3]-labeled pre-rRNA and rRNA was resolved on 1.2% agarose gels for high-molecular-weight RNAs and 8% acrylamide-8 M urea gels for low-molecular-weight RNAs. RNA was transferred to Hybond-N$^+$ Nylon membranes (Amersham), dried, and exposed to X-ray film for 10 days at −80 °C with an intensifying screen.

ACKNOWLEDGMENTS

M.O. was the recipient of a Darwin Trust Fellowship and A.L. was the recipient of a Studentship from the Croucher Foundation of Hong Kong. A.L. and D.T. are Wellcome Trust Principal Fellows. This work was supported by the Wellcome Trust.

Received December 10, 2001; returned for revision January 3, 2002; revised manuscript received February 22, 2002

REFERENCES

Allende ML, Amsterdam A, Becker T, Kawakami K, Gaiano N, Hopkins N. 1996. Insertional mutagenesis in Zebrafish identifies two novel genes, Pescadillo and dead eye, essential for embryonic development. *Genes & Dev 10*:3141–3155.

Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D. 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J 18*:5399–5410.

Anderson J, Lyon CE, Fox A, Leung AKL, Lam YW, Steen H, Mann M, Lamond AI. 2002. Directed proteomic analysis of the human nucleolus. *Current Biology 12*:1229–1231.

Baßler J, Grandi P, Gadal O, Leßmann T, Petfalski E, Tollervey D, Lechner J, Hurt E. 2001. Identification of a 60S pre-ribosomal particle that is closely linked to nuclear export. *Mol Cell 8*:517–529.

Beltrame M, Tollervey D. 1992. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J 11*:1531–1542.

Bergès T, Petfalski E, Tollervey D, Hurt EC. 1994. Synthetic lethality with fibrillarin identifies NOP77p, a nucleolar protein required for pre-rRNA processing and modification. *EMBO J 13*:3136–3148.

Biggiogera M, Burki K, Kaufmann SH, Shaper JH, Gas N, Amalric F, Fakan S. 1990. Nucleolar distribution of proteins B23 and nucleolin in mouse preimplantation embryos as visualized by immunoelectron microscopy. *Development 110*:1263–1270.

Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV. 1997. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J 11*:68–76.

de la Cruz J, Kressler D, Tollervey D, Linder P. 1998. Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3′ end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J 17*:1128–1140.

Ding DQ, Tomita Y, Yamamoto A, Chikashige Y, Haraguchi T, Hiraoka Y. 2000. Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library. *Genes Cells 5*:169–190.

Fatica A, Cronshaw AD, Dlakic M, Tollervey D. 2002. Ssf1p prevents premature processing of an early pre-60S ribosomal particle. *Mol Cell 9*:341–351.

Fuchs SY, Adler V, Buschmann T, Wu X, Ronai Z. 1998. Mdm2 association with p53 targets its ubiquitination. *Oncogene 17*:2543–2547.

Gadal O, Strauss D, Braspenning J, Hoepfner D, Petfalski E, Philippsen P, Tollervey D, Hurt E. 2001a. A nuclear AAA-type ATPase (Rix7p) is required for biogenesis and nuclear export of 60S ribosomal subunits. *EMBO J 20*:3695–3704.

Gadal O, Strauss D, Kessl J, Trumpower B, Tollervey D, Hurt E. 2001b. Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol Cell Biol 21*:3405–3415.

Geerlings TH, Vos JC, Raue HA. 2000. The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 5′ → 3′ exonucleases. *RNA 6*:1698–1703.

Geyer RK, Yu ZK, Maki CG. 2000. The MDM2 RING-finger domain is required to promote p53 nuclear export. *Nat Cell Biol 2*:569–573.

Grandi P, Doyl V, Hurt EC. 1993. Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. *EMBO J 12*:3061–3071.

Haque J, Boger S, Li J, Duncan SA. 2000. The murine Pes1 gene encodes a nuclear protein containing a BRCT domain. *Genomics 70*:201–210.

Harnpicharnchai P, Jakovljevic J, Horsey E, Miles T, Roman J, Rout M, Meagher D, Imai B, Guo Y, Brame CJ, Shabanowitz J, Hunt DF, Woolford JL. 2001. Composition and functional characterization of yeast 66s ribosome assembly intermediates. *Mol Cell 8*:505–515.

Henry Y, Wood H, Morrissey JP, Petfalski E, Kearsey S, Tollervey D. 1994. The 5′ end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J 13*:2452–2463.

Ho JH, Kallstrom G, Johnson AW. 2000. Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *J Cell Biol 151*:1057–1066.

Hurt E, Hannus S, Schmelzl B, Lau D, Tollervey D, Simos G. 1999. A novel in vivo assay reveals inhibition of ribosomal nuclear export in ran-cycle and nucleoporin mutants. *J Cell Biol 144*:389–401.

Kinoshita Y, Jarell AD, Flaman JM, Foltz G, Schuster J, Sopher BL, Irvin DK, Kanning K, Kornblum HI, Nelson PS, Hieter P, Morrison RS. 2001. Pescadillo, a novel cell cycle regulatory protein abnormally expressed in malignant cells. *J Biol Chem 276*:6656–6665.

Kressler D, Linder P, de La Cruz J. 1999. Protein *trans*-acting factors involved in ribosome biogenesis in *Saccharomyces cerevisiae*. *Mol Cell Biol 19*:7897–7912.

Lafontaine D, Tollervey D. 1996. One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res 24*:3469–3472.

Lafontaine DL, Tollervey D. 2001. The function and synthesis of ribosomes. *Nat Rev Mol Cell Biol 2*:514–520.

Leger-Silvestre I, Noaillac-Depeyre J, Faubladier M, Gas N. 1997. Structural and functional analysis of the nucleolus of the fission yeast Schizosaccharomyces pombe. *Eur J Cell Biol 72*:13–23.

Leger-Silvestre I, Trumtel S, Noaillac-Depeyre J, Gas N. 1999. Functional compartmentalization of the nucleus in the budding yeast *Saccharomyces cerevisiae*. *Chromosoma 108*:103–113.

Lyon CE, Lamond AI. 2000. The nucleolus. *Curr Biol 10*:R323.

Milkereit P, Gadal O, Podtelejnikov A, Trumtel S, Gas N, Petfalski E, Tollervey D, Mann M, Hurt E, Tschochner H. 2001. Maturation and intranuclear transport of pre-ribosomes requires Noc proteins. *Cell 105*:499–509.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol 17*:1030–1032.

Sakumoto N, Yamashita H, Mukai Y, Kaneko Y, Harashima S. 2001. Dual-specificity protein phosphatase Yvh1p, which is required for vegetative growth and sporulation, interacts with yeast Pescadillo homolog in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun 289*:608–615.

Savkur RS, Olson MO. 1998. Preferential cleavage in pre-ribosomal RNA byprotein B23 endoribonuclease. *Nucleic Acids Res 26*: 4508–4515.

Scheer U, Hock R. 1999. Structure and function of the nucleolus. *Curr Opin Cell Biol 11*:385–390.

Schimmang T, Tollervey D, Kern H, Frank R, Hurt EC. 1989. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J 8*:4015–4024.

Schmitt ME, Brown TA, Trumpower BL. 1990. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res 18*:3091–3092.

Shaw PJ, Jordan EG. 1995. The nucleolus. *Annu Rev Cell Dev Biol 11*:93–121.

Shou W, Sakamoto KM, Keener J, Morimoto KW, Traverso EE, Azzam R, Hoppe GJ, Feldman RM, DeModena J, Moazed D, Charbonneau H, Nomura M, Deshaies RJ. 2001. Net1 stimulates RNA polymerase I transcription and regulates nucleolar structure independently of controlling mitotic exit. *Mol Cell 8*:45–55.

Shou W, Seol JH, Shevchenko A, Baskerville C, Moazed D, Chen ZW, Jang J, Shevchenko A, Charbonneau H, Deshaies RJ. 1999. Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. *Cell 97*:233–244.

Stage-Zimmermann T, Schmidt U, Silver PA. 2000. Factors affecting nuclear export of the 60S ribosomal subunit in vivo. *Mol Biol Cell 11*:3777–3789.

Tao W, Levine AJ. 1999. P19(ARF) stabilizes p53 by blocking nucleo-cytoplasmic shuttling of Mdm2. *Proc Natl Acad Sci USA 96*: 6937–6941.

Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J 6*:4169–4175.

Tollervey D, Lehtonen H, Jansen R, Kern H, Hurt EC. 1993. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell 72*:443–457.

Tollervey D, Mattaj IW. 1987. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J 6*:469–476.

Venema J, Tollervey D. 1999. Ribosome synthesis in *Saccharomyces cerevisiae*. *Ann Rev Gen 33*:261–311.

Warner JR. 2001. Nascent ribosomes. *Cell 107*:133–136.

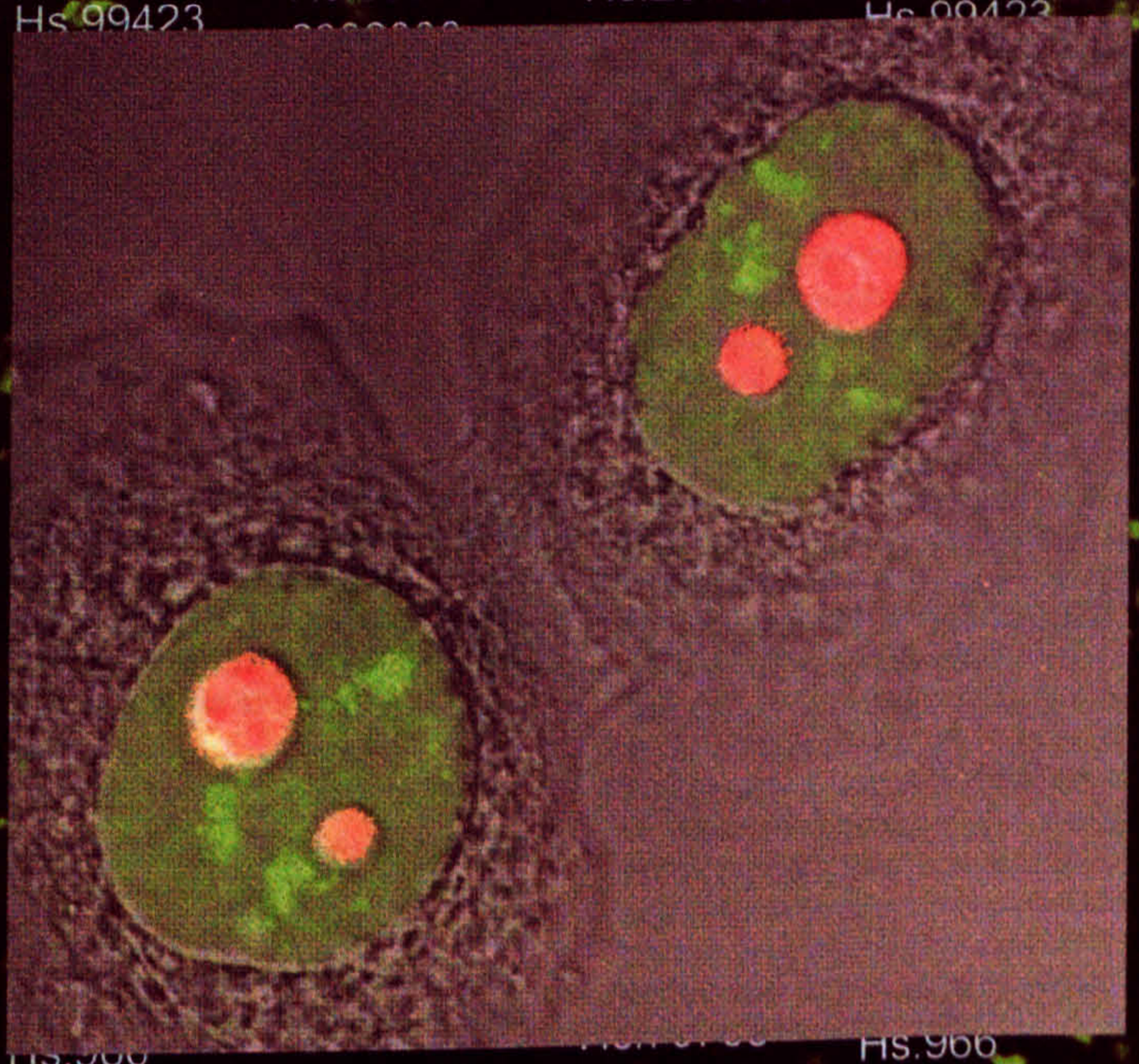
Weber JD, Taylor LJ, Roussel MF, Sherr CJ, Bar-Sagi D. 1999. Nucleolar Arf sequesters Mdm2 and activates p53. *Nat Cell Biol 1*:20–26.

Zhang Y, Xiong Y. 1999. Mutations in human ARF exon 2 disrupt its nucleolar localization and impair its ability to block nuclear export of MDM2 and p53. *Mol Cell 3*:579–591.

Current Biology

Volume 12 Number 1

January 8, 2002



Human Nucleolar Proteome

Current Biology, Vol. 12, 1–11, January 8, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0960-9822(01)00650-9

Directed Proteomic Analysis of the Human Nucleolus

Jens S. Andersen,[1] Carol E. Lyon,[2] Archa H. Fox,[2] Anthony K.L. Leung,[2] Yun Wah Lam,[2] Hanno Steen,[1] Matthias Mann,[1,3,4] and Angus I. Lamond[2,4]
[1]Department of Biochemistry and Molecular Biology
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark
[2]Wellcome Trust Biocentre
MSI/WTB Complex
University of Dundee
Dundee DD1 4HN
United Kingdom
[3]MDS Proteomics
Staermosegaadsvej 6
5230 Odense M
Denmark

Summary

Background: The nucleolus is a subnuclear organelle containing the ribosomal RNA gene clusters and ribosome biogenesis factors. Recent studies suggest it may also have roles in RNA transport, RNA modification, and cell cycle regulation. Despite over 150 years of research into nucleoli, many aspects of their structure and function remain uncharacterized.

Results: We report a proteomic analysis of human nucleoli. Using a combination of mass spectrometry (MS) and sequence database searches, including online analysis of the draft human genome sequence, 271 proteins were identified. Over 30% of the nucleolar proteins were encoded by novel or uncharacterized genes, while the known proteins included several unexpected factors with no previously known nucleolar functions. MS analysis of nucleoli isolated from HeLa cells in which transcription had been inhibited showed that a subset of proteins was enriched. These data highlight the dynamic nature of the nucleolar proteome and show that proteins can either associate with nucleoli transiently or accumulate only under specific metabolic conditions.

Conclusions: This extensive proteomic analysis shows that nucleoli have a surprisingly large protein complexity. The many novel factors and separate classes of proteins identified support the view that the nucleolus may perform additional functions beyond its known role in ribosome subunit biogenesis. The data also show that the protein composition of nucleoli is not static and can alter significantly in response to the metabolic state of the cell.

[4]Correspondence: a.i.lamond@dundee.ac.uk (A.I.L.), mann@bmb.sdu.dk (M.M.)

Introduction

Many studies have shown that the cell nucleus contains distinct and often dynamic compartments [1–4]. Chromosome territories are separated by the interchromatin space, which contains discrete nuclear bodies, including nucleoli, Cajal bodies, speckles, gems, and promyelocytic leukemia (PML) bodies that have mainly been characterized at the morphological level [5]. Nuclear bodies contain complexes of proteins and/or RNPs but are not separated by membranes from the surrounding nucleoplasm. In vivo experiments using fluorescence photobleaching (FRAP and FLIP) techniques have shown that their protein components move by passive diffusion and are in continuous flux between nuclear compartments [6, 7]. Self-organization has been suggested as a mechanism for nuclear body formation and maintenance [8]. In addition, nuclear bodies themselves can move in the nucleoplasm. For example, the mobility of Cajal bodies [9] may also involve diffusion, although active and directed mechanisms could also play a role [10].

In most cases, the biological roles of nuclear bodies are still not well understood, and their molecular constituents have not been comprehensively identified. Recent advances in MS techniques, coupled with the accumulation of genomic and EST DNA sequence databases, have facilitated the global characterization of protein complexes and organelles [11–14]. In the case of the nucleus, a recent study has applied MS to analyze the protein composition of fractions from mouse liver cells enriched in interchromatin granule clusters (IGCs) [15]. Previous work from our own laboratories and from other groups has employed proteomic methods to characterize the protein composition of a variety of multiprotein complexes and subcellular organelles from yeast to human, including U1 snRNPs [16], spliceosomes [17], and nuclear pore complexes [18].

The best-studied example of a membrane-free nuclear subdomain is the nucleolus, a "cellular factory" in which 28S, 18S, and 5.8S ribosomal RNAs (rRNAs) are transcribed and together with 5S rRNA are processed and assembled into the ribosome subunits. The nucleolus is a dynamic structure that disassembles and reforms during each cell cycle around the rRNA gene clusters [19]. Within the nucleolus, three distinct subcompartments have been described based on their morphology in the electron microscope. These correspond to the fibrillar centers (FC), dense fibrillar components (DFC), and granular components (GC). The current consensus view is that rDNA transcription is restricted to the periphery of the FC, while transient accumulation, modification, and processing of primary rRNA transcripts occurs in the DFC, and later processing and rRNA assembly into ribosomal subunits occurs in the GC [20–23].

While it is clear that the major role of the nucleolus is in ribosome subunit biogenesis, it is interesting that recent studies suggest that there may be additional functions for the nucleolus [24–26]. For example, it may also be a site for the biogenesis and/or maturation of

other ribonucleoprotein machines, including the signal recognition particle [27], the spliceosomal small nuclear RNPs [28], and telomerase [29]. The nucleolus may also participate in processing or export of some mRNAs and tRNAs [26, 30] and can control the activities of specific regulatory factors by a sequestration mechanism [31]. Moreover, association between the nucleolus and other nuclear bodies, such as the perinucleolar compartment (PNC) [32] and the Cajal body [33], also raises the possibility that nucleoli have additional functions.

Here we present an analysis of the proteome of purified human nucleoli, using MS and computer searches in sequence databases, including the draft human genome. Over 30% of the 271 nucleolar proteins identified are encoded by novel human genes, and a subset of factors are shown to accumulate in nucleoli specifically when transcription is inhibited.

Results

Isolation of Nucleoli

It was essential to have highly purified preparations of nucleoli as a source material for direct protein analysis. The inherent density of nucleoli [34] facilitated their isolation from cultured cells, using a combination of sonication and sucrose density centrifugation (see Experimental Procedures). Thus, nucleoli were isolated from HeLa cell nuclei, and their purity and integrity were analyzed using both light and electron microscopy and protein blotting (Figure 1). As judged by light microscopy, over 95% of the particles detected also stained positive with the RNA-specific dye Pyronin Y and were immunolabeled with antibodies to nucleolar proteins (Figure 1A). The isolated nucleoli were embedded in resin, sectioned, and examined in the electron microscope (Figure 1B), showing that they had typical nucleolar morphology, as seen in sections through intact nuclei (Figure 1B, compare nucleoli in situ [1Bi] with isolated nucleoli [panels 1Bii–1Biv]). At the highest magnification, the internal nucleolar substructures, corresponding to the FC, DFC, and GC, are clearly visible in sections of the isolated nucleoli (Figure 1Biv). To address the purity of the isolated nucleoli using a biochemical assay, protein samples from (1) unfractionated nuclear extract, (2) nucleoplasmic fraction, and (3) purified nucleoli were separated by SDS PAGE, transferred to a nitrocellulose membrane, and probed with antibodies specific for either nucleolin, fibrillarin, Lamin B, or NUP 62 (Figure 1C). This shows that both nucleolar proteins tested, i.e., nucleolin and fibrillarin, are highly enriched in the nucleolar fraction (compare lanes 2 and 3). In contrast, two nuclear proteins that do not accumulate in nucleoli, i.e., lamin B and NUP62, are not detected in the purified nucleolar fraction.

To evaluate the suitability of these purified samples for large-scale MS characterization, a preliminary MS analysis was conducted following separation of the purified nucleolar proteins by one-dimensional (1D) SDS PAGE. Gel lanes were completely sliced, in-gel digested with trypsin, and the resulting peptides analyzed automatically by high mass-accuracy peptide-mass mapping using MALDI-MS (Figure 2 and data not shown). These data were screened against a human-specific nonredundant protein database (nrdb). The resulting list of 80 proteins included many known nucleolar proteins without obvious major protein contaminants (see below). We conclude that the isolation procedure yields intact preparations of HeLa cell nucleoli that are sufficiently pure for use in proteomic analysis.

Identification and Characterization of Nucleolar Proteins

The purified nucleoli were used in a large-scale MS analysis, involving the identification of proteins separated by both 1D and 2D SDS PAGE (Figure 2; Table S1 [Table S1 is included as a special reprint of this article and as Supplementary Material available online]). A variety of different gel/buffer systems, which each have advantages for resolving different types of proteins (for example, compare Figures 2A and 2B), were used to ensure maximal coverage of nucleolar proteins. However, it was not necessary to resolve every individual protein component prior to MS analysis, because nanoelectrospray (nanoES) tandem MS could identify separate proteins even in complex mixtures, including up to eight different factors (Figure 3A). The majority of nucleolar proteins could therefore be identified by analyzing proteins separated using the different 1D SDS PAGE systems followed by nanoES sequencing of proteins in individual gel slices.

To facilitate the analysis of the complex range of nucleolar proteins, a layered analytical strategy [35] was refined to allow online interrogation of complex peptide mixtures and to identify proteins directly in genome sequences (Figure 3). Peptide sequence tags [36] assigned from fragmentation spectra of peptide signals were searched against various sequence databases in "real time" during data acquisition. The sequences of other expected peptides in the mixture were then identified on the basis of already identified proteins. These peptides were labeled and excluded from further sequencing, thereby sequentially reducing the complexity of the peptide mixture. This approach, termed "directed sequencing," proved to be highly efficient at allocating sequencing time to the identification of the largest number of proteins from each sample (Figure 3B). Unambiguous sequence identification was established by directed sequencing of a minimum of two peptides for each protein, followed by matching expected and measured peptide fragment ions. Directed sequencing of key peptides was particularly useful in (1) the identification of low abundance proteins, (2) the identification of proteins represented by limited sequence information, e.g., EST's, and (3) to differentiate between protein isoforms, for example, PSP1 where sequencing of the peptide ALESAVGESE PAAAAAMALALAGEPAPPAPAPPEDHPDEEMGFTI DIK was required to confirm the N-terminal extended form [37].

The 1D PAGE analyses showed that the majority of nucleolar proteins were in the size range ~25–100 kDa. 2D gel electrophoresis was used to expand the separation of proteins in this size range (Figure 2C). Individual spots from the 2D gels were detected as described in Experimental Procedures, excised, and in-gel digested with trypsin prior to MS analysis. These data confirmed

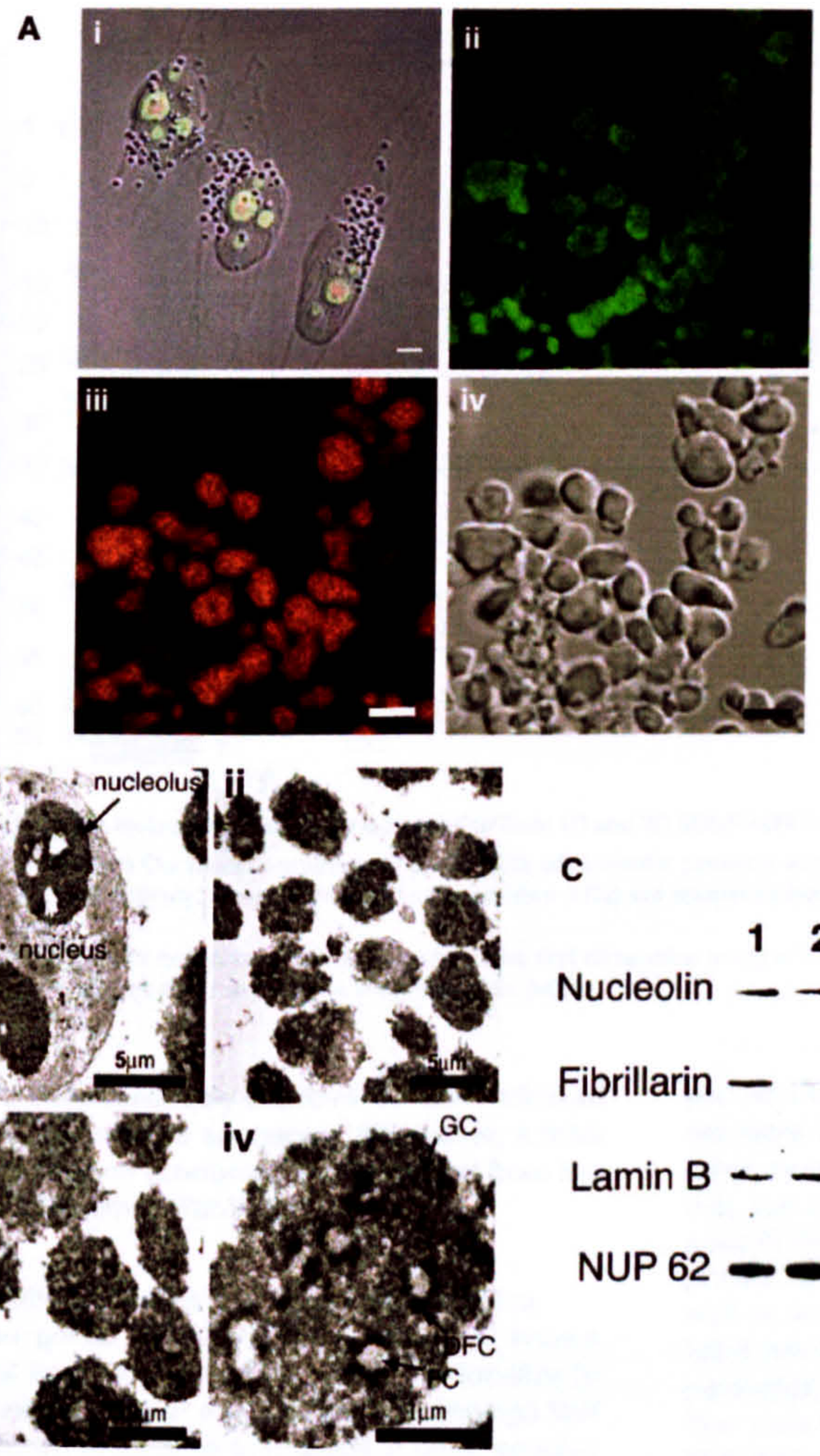


Figure 1. Nucleoli Can Be Isolated from HeLa Cells in Large Quantities and with High Purity

The purity of the isolated nucleolar preparations was monitored using several methods. (A) Light microscopy can demonstrate the purity of purified HeLa nucleoli. Panel (Ai) shows an intact HeLa cell stained with two nucleolar-specific labels: Pyronin Y (red) and nucleolin antibody (green). These reagents were also used to stain purified nucleoli (nucleolin in panel [Aii] and Pyronin Y in panel [Aiii]). DIC imaging of the purified material (Aiv) also shows that it does not contain significant amounts of nonnucleolar particles. (B) TEM images of thin sections of a HeLa cell nucleus with two nucleoli in situ, marked by arrows (Bi). Panels (Bii)–(Biv) show TEM images of sections of isolated nucleoli in increasing magnification. These clearly show that the isolated nucleoli are morphologically intact; they contain clearly defined granular component (GC), dense fibrillar component (DFC), and fibrillar centers (FC). (C) Proteins from preparations of sonicated whole nuclei, nucleoplasm, and nucleoli were separated by gel electrophoresis, transfered to nitrocellulose, and immunolabeled with anti-nucleolin, anti-fibrillarin, anti-lamin B, and anti-Nup 62. The nucleolar fraction is enriched in nucleolin and fibrillarin but not in the nonnucleolar proteins.

the identification of many of the factors detected in the 1D gel systems and identified 20 additional proteins (Table S1). However, many basic proteins identified by 1D PAGE analyses were not resolved in the 2D gel system.

The peptide sequence tag search algorithm [36] was adapted to allow protein identification directly in the draft genome sequence [38]. Protein identity was confirmed by detecting two or more peptides matching within either a single exon or within different exons confined to the same genomic region. Examples of direct MS gene identification in the unannotated human genome are shown in Figures 3C–3E. In the first example, a peptide sequence tag (underlined in Figure 3D) matched exclusively a novel exon located in the genome sequence within chromosome 16. Directed sequencing of three additional peptides (underlined in Figure 3D), predicted from the same exon, confirmed the identification. The coding sequence of this unusual large exon, encoding a putative exoribonuclease, had a calculated M_r of 25.993 kDa, consistent with the observed M_r of 26 kDa for the cognate gel band. The second example illustrates genome searching combined with gene prediction (Figure 3E). In this case, three peptides that were identified all mapped to a short region of the genome, spanning ~100 kb within chromosome 12. This region was therefore used as a constraint for further analysis. Sequencing of additional peptides that either lay within or spanned exons from this region allowed us to refine the predicted gene structure at this locus, including the identification of an additional exon. Importantly, this MS analysis also identified the N-acetylated N-terminal peptide that includes the initiating methionine. A total of 14 novel proteins that copurify with nucleoli were identified directly by analysis of the human genome sequence (Table S1). In other cases, peptides identified and matched to the genome sequence allowed us to revise the structures of predicted genes retrieved from a nonredundant database (for example, GCN1L1, SAZD, and BING4). The above examples show how the MS data, combined with direct analysis of the human genome sequence, facilitate the identification of novel nucleolar

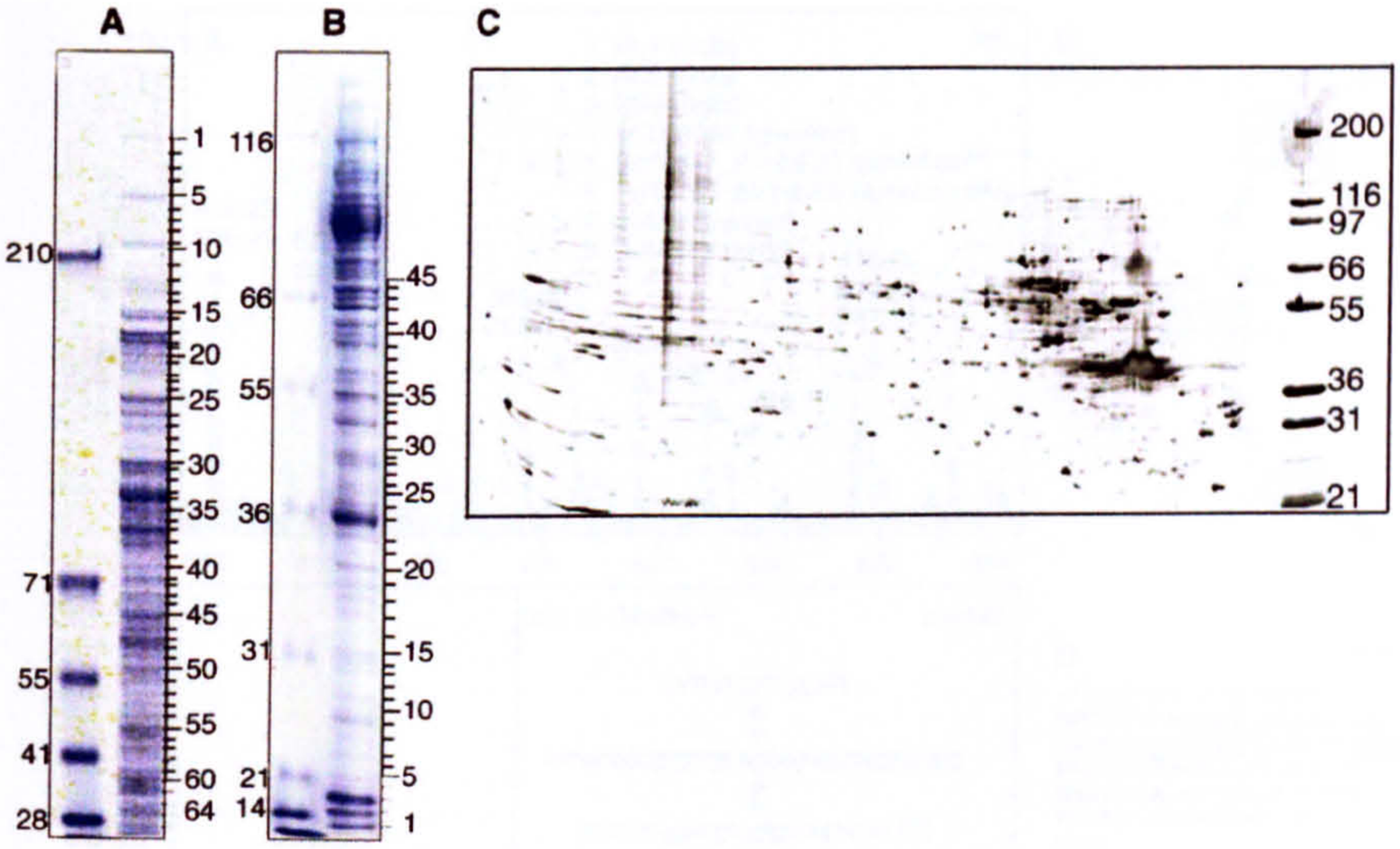


Figure 2. Proteins from Isolated Nucleoli Were Resolved by Both 1D and 2D SDS/PAGE to Maximize the Coverage of Nucleolar-Associated Proteins

(A and B) Representative Coomassie-stained 1D SDS/PAGE of nucleolar proteins separated utilizing either a tris-glycine (A) or a tris-acetate buffer system (B), respectively. Molecular weight size markers (kDa) are shown to the left of each gel, and rulers used for excising gel slices are shown at right.

(C) Silver-stained 2D gel of nucleolar samples focused in the first dimension using a 3-10 strip then subsequently electrophoresed on a 12.5% SDS gel. Molecular weight size markers are shown at right (kDa).

proteins and simultaneously improve the annotation of the draft human genome sequence. Altogether, a total of 271 separate gene products were identified from the nucleolar preparations (Table S1).

Characterization of Identified Nucleolar Proteins

Many of the genes identified above encoded known proteins that had previously been shown to localize to nucleoli in human cells or else had close homologs that had been shown to localize to nucleoli in other species (Table S1). However, over 30% of the genes encoded novel or uncharacterized proteins whose localization was unknown. In addition, some of the known gene products identified had not been reported as nucleolar factors. To address whether these classes of proteins are genuine nucleolar components in vivo, we adopted the strategy, as in our previous study of the spliceosome [17], of tagging them with fluorescent proteins and examining their localization in the fluorescence microscope following transient expression in cultured cells (Figure 4). Due to the large number of proteins involved, it was beyond the scope of this study to isolate and tag full-length cDNA clones for every gene identified. Therefore, we selected for analysis 18 examples of proteins that were either novel, or not known to be nucleolar, isolated cDNA clones and fused them to yellow fluorescent protein (YFP) at their amino termini in the pEYFPC1 expression vector (see Experimental Procedures). The proteins selected were chosen to include a range of sizes, pI values, and motifs. Following transient transfection and expression in HeLa cells, 15 of the fusion proteins localized to nucleoli, showing a variety of different patterns (Figure 4 and data not shown). Some of the proteins localized exclusively to nucleoli, whereas others were detected in nucleoli but also accumulated at other nuclear and/or cytoplasmic structures. For example, the human Pescadillo homolog (PES1) localized specifically to nucleoli (Figure 4A), whereas the PWP1 protein showed extensive cytoplasmic localization as well as accumulating in nucleoli (Figure 4D). Some proteins were detected within distinct subregions of the nucleolus, such as the novel protein SAZD (Figure 4E). The variety of localization patterns observed with the YFP-fusion proteins is consistent with the wide spectrum of nucleolar labeling patterns detected previously for known nucleolar proteins. During the course of our analyses, several recent studies have independently reported the nucleolar localization of some of the proteins we have identified by MS (for example, NOH61 [39] and Bop1 [40]). As discussed below, the YFP-tagging data support the view that the majority of the novel proteins detected here by MS can interact with nucleoli in vivo. Some or all of the tagged proteins that could not be confirmed here as nucleolar may also interact with nucleoli in vivo (see Discussion).

An overview of the separate classes of proteins identified in the purified nucleoli reported in Table S1 is presented in the form of a pie chart in Figure 5. The largest single category (~31%) represents the group of novel and previously uncharacterized factors, including those whose function cannot be readily inferred based upon recognizable homology to known proteins. The next largest category (24%) corresponds to proteins that bind to nucleic acid and/or to nucleotides, consistent with the known role of the nucleolus in ribosome biogenesis and the presence in nucleoli of large amounts of rRNA and the rRNA repeat gene clusters. Other classes listed

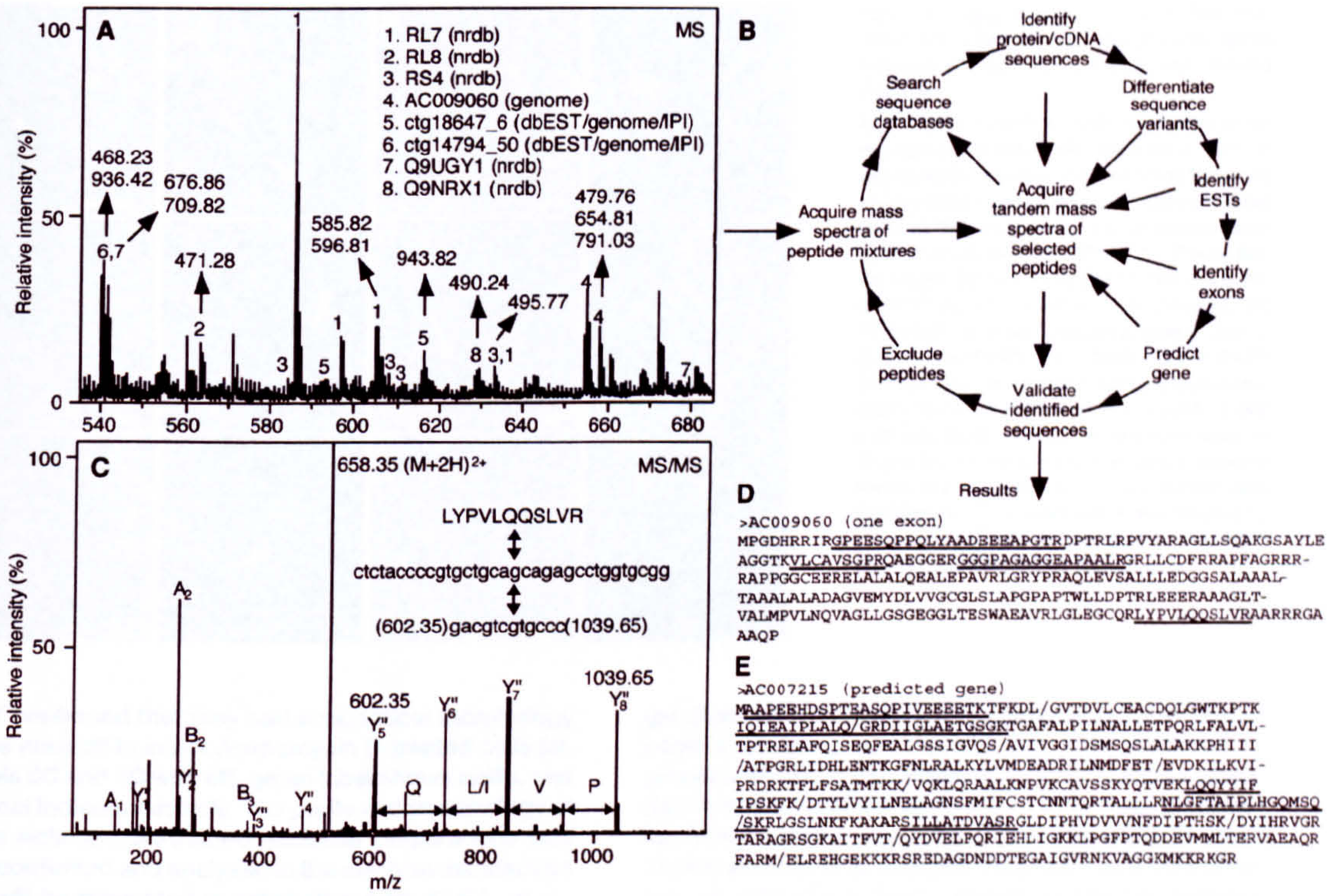


Figure 3. Identification of Proteins from Isolated Nucleoli

The gel lanes/spots from the 1D and 2D gels of nucleolar proteins (see Figure 2) were cut into slices, enzymatically digested, and the resulting peptide mixtures analyzed by MALDI and nanoES MS. (A) Part of nanoES mass spectrum of tryptic peptides from gel slice 18P. Marked peaks (1) were fragmented and validated identification of RL7 by MALDI-MS peptide mapping. Tandem MS of unexplained peptide signals, followed by peptide sequence tag database searches, resulted in identification of seven additional proteins. Database searches in real time allowed unambiguous identification by directed sequencing of at least two peptides for each of the retrieved sequences. (B) Flow chart illustrating the directed sequencing approach to identify and validate a maximum number of proteins from a peptide mixture. (C–E) Examples of identifying proteins in the human genome. (C) Fragment ion mass spectrum of the peak at m/z 558.35. A peptide sequence tag was assembled from the series of C-terminal fragment ions. The amino acid sequence of the search string was translated into the corresponding degenerated nucleotide sequence. Potential hits in the forward or reverse direction of the genome data were checked as to whether they coded for the amino acid sequence defined by the peptide sequence tag. Additional N- and C-terminal fragment ions (B and Y″ ions) matched the retrieved peptide LYPVLQQSLVR. (D) Amino acid sequence translated from the identified exon. Underlined peptides were fragmented to confirm identification and to narrow down the probable terminal splice sites. (E) Amino acid sequence translated from a gene predicted from genomic data using the peptides IGLAETGSGK, LQQYYIFIPSK, and ILLATDVASR identified in slice 36P as coding constraints. Underlined peptides were sequenced, including the N-acetylated N-terminal peptide and peptides spanning exon boundaries that partially confirmed the gene prediction.

also include families of RNA or nucleotide binding proteins whose functions are more clearly defined. For example, the DEAD box motif that is characteristic of the superfamily of RNA-dependent ATPases and helicases [41] is found in 8% of the total nucleolar proteins, including 3% of the novel factors. The ribosomal proteins and other translation factors comprise a similar fraction of nucleolar proteins as do the RNA modifying enzymes (including snoRNP proteins) and chaperones. The category listed as "others" includes proteins with a wide range of known motifs and disparate functions. Collectively, these data illustrate the surprisingly high complexity of the nucleolar proteome.

Actinomycin D Treatment Changes the Nucleolar Proteome

We analyzed whether the nucleolar proteome was affected by treatment of HeLa cells with Actinomycin D, which is known to affect the structure of the nucleus, the localization pattern of nuclear proteins, and the integrity of the nucleolus [42]. The efficacy of Actinomycin D treatment was confirmed by double labeling treated and control HeLa cells with both pyronin Y and anti-coilin antibodies (Figures 6A and 6B). As expected [43, 44], Actinomycin D-treated cells showed a relocalization of coilin to form caps at the nucleolar periphery (cf. panels 6A and 6B, green label shows coilin and red foci indicate nucleoli). Additional experiments to monitor the incorporation of Br-UTP before and after Actinomycin D treatment indicated that both RNA polymerase I and II activities had been inhibited by the drug treatment (data not shown). Therefore, we isolated quantitative preparations of nucleoli from Actinomycin-treated cells to allow for direct MS analysis of their protein composition (Figure 6). Double labeling of the nucleoli isolated from Actinomycin-treated cells with anti-coilin antibodies and pyro-

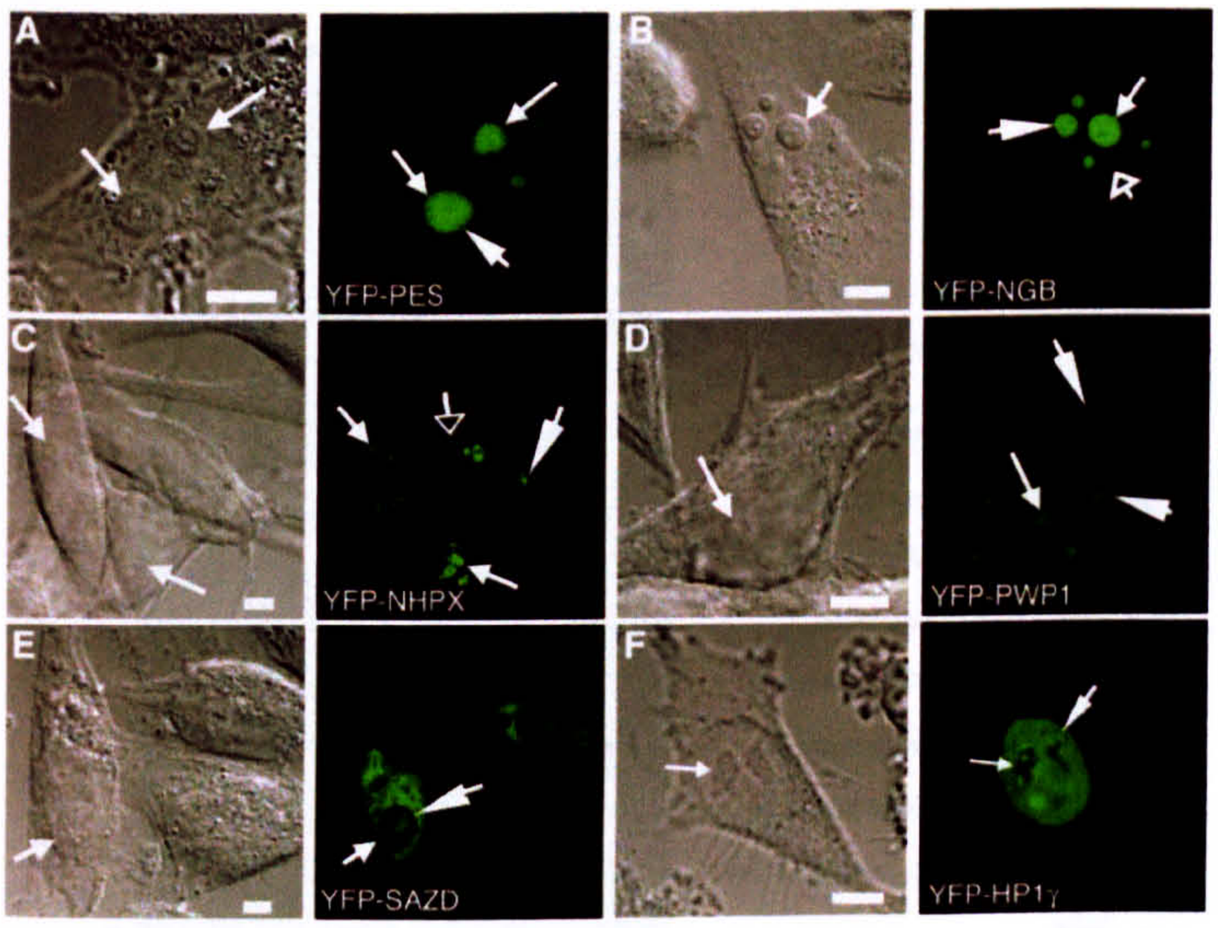


Figure 4. Tagged Nucleolar Candidate Proteins, Identified by MS, Accumulate within Nucleoli in HeLa Cells, Albeit with Varying Patterns

Each panel shows a confocal fluorescence micrograph of HeLa cells expressing YFP fusion proteins, with the corresponding Nomarski image shown on the left. Constructs expressed were as follows (names and accession numbers in brackets): (A) YFP-PES1 (Pescadillo, Hs.13501), (B) YFP-NGB (GTP binding protein, Hs.215766), (C) YFP-NHPX (Hs.182255), (D) YFP-PWP1 (nuclear phosphoprotein similar to *S. cerevisiae* PWP1, Hs.172589), (E) YFP-SAZD (Hs.114416), and (F) YFP-HP1γ (Heterochromatin Protein 1 γ isoform, Hs.278554). Small arrows indicate nucleoli, large arrowheads indicate the localization of the fusion proteins within the nucleoli, and open arrowheads highlight small nuclear bodies also labeled by the fusion proteins (in the case of YFP-NGB and YFP-NHPX). Scale bars, 5 μm.

nin Y confirmed that they had an identical morphology to the nucleoli in intact Actinomycin D-treated cells (cf. panels 6C and 6D with 6B, green label shows coilin, and red foci indicate nucleoli). The purity and morphology of these Actinomycin-treated nucleolar preparations was also confirmed and analyzed in the electron microscope and will be presented in detail elsewhere (C.E.L. et al., unpublished data).

Proteins from nucleoli isolated from both control and Actinomycin D-treated HeLa cells were separated by 1D SDS PAGE and stained with Coomassie dye (Figure 6G). The total proteomes are similar for both control and Actinomycin D-treated nucleoli as seen by SDS-PAGE (Figure 6G) and by additional MS analysis (data not shown). However, there were some protein bands whose intensity in the Actinomycin-treated nucleoli was increased relative to the control nucleoli (Figure 6G, cf. Lanes 1 and 2, arrows). These bands were excised, in-gel digested with trypsin, and analyzed by both MALDI peptide mapping and nanoES MS. This identified 11 proteins as major candidates for factors whose abundance in the nucleolar proteome was increased following Actinomycin D treatment (Figure 6G, proteins listed beside arrows). This included the protein p80 coilin (running at ~66 kDa in both extracts in this gel system), which was shown by immunocytochemical analysis to have increased association with nucleoli following Actinomycin D treatment (see also Figures 6A–6D). Interestingly, the other ten proteins identified were members of three separate protein families, i.e., DEAD box proteins (DDX9, p72, and p68), hnRNP proteins (hnRNPs K, G, and A2/B1), and a group of related RNA binding proteins (PSF, PSP2/CoAA, PSP1, and p54/nrb). We therefore selected a number of proteins from two of these families (i.e., p68, p72, p54/nrb, PSP1, and PSP2/CoAA), tagged them with YFP, and compared their localization in both control and Actinomycin-treated cells (Figures 6E and 6F show one example where the green signal shows YFP-p68 and the red channel Pyronin Y; see also [37] and other data not shown). Surprisingly, all the proteins analyzed were predominantly nucleoplasmic but relocated to the nucleolar periphery following Actinomycin D treatment (a detailed analysis of the localization and behavior of PSP1 is reported in [37]). Furthermore, consistent with this MS identification, a recent study has reported that GFP-PSF also relocalizes to the nucleolar periphery following Actinomycin D treatment [45].

In summary, we conclude that metabolic perturbations, as induced here by Actinomycin D treatment, can induce changes in the relative abundance of a subset of the nucleolar proteome.

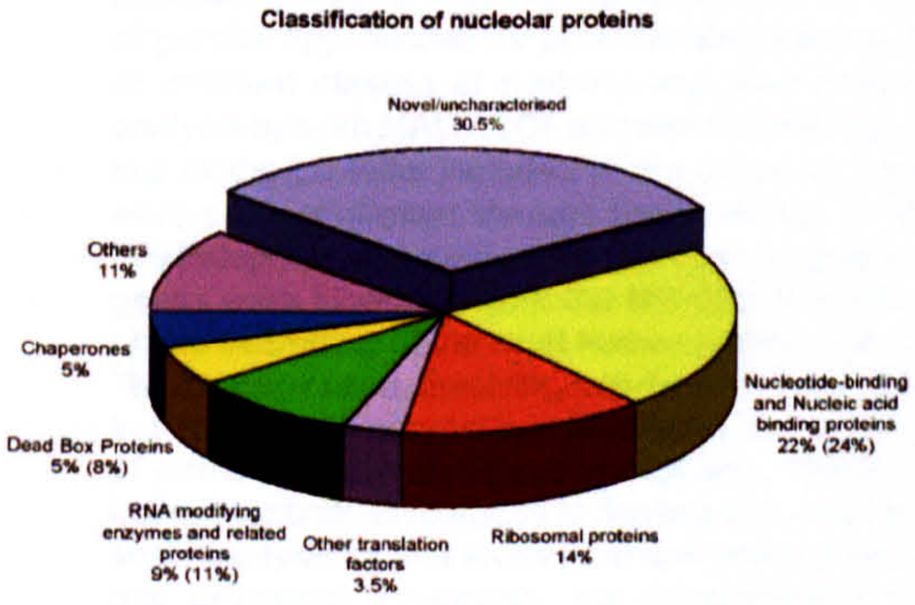


Figure 5. Distribution of Conserved Motifs and Putative Functional Categories of the Identified Proteins

The proteins from Table S1 were divided into groups as indicated above. The percentages in brackets indicate the number of proteins, including those classed as novel/uncharacterized, that contain the motif in question.

Discussion

We have reported here a characterization of the proteome of nucleoli isolated from cultured human cells. This study represents the largest proteomic analysis

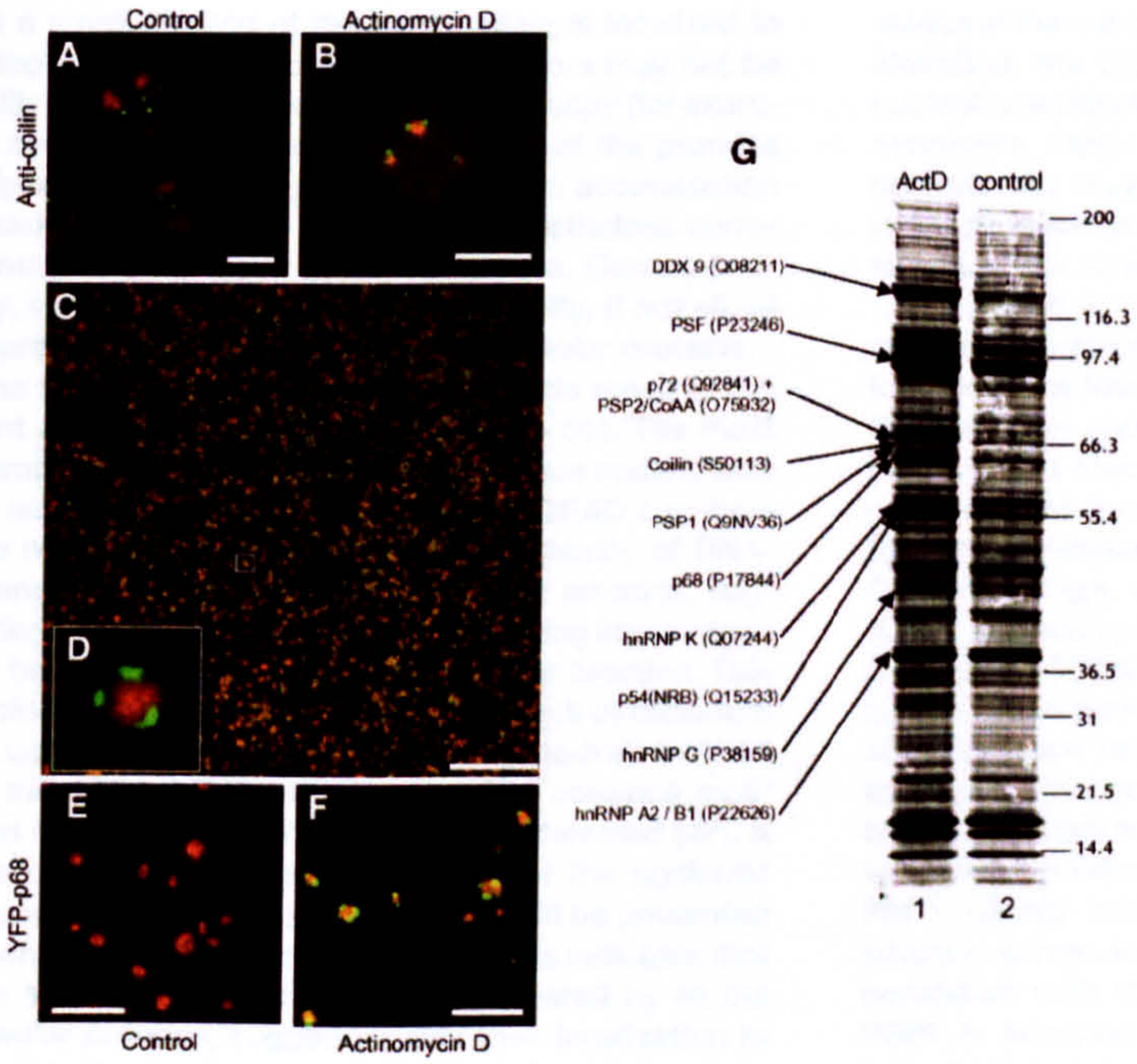


Figure 6. Changes in Nucleolar Structure and Composition after Actinomycin D Treatment of HeLa Cells

Following incubation of HeLa cells with Actinomycin D (see Experimental Procedures), coilin (as detected with anti-coilin antiserum, green) relocalized from Cajal bodies, as seen in untreated cells (A), to caps, labeled by Pyronin Y (red, [B]). These perinucleolar coilin caps remained detectable in nucleoli isolated from Actinomycin D-treated cells (C). (D) A magnified view of one of the isolated nucleoli. (G) Proteins from nucleoli isolated from Actinomycin D-treated cells (lane 1) and the untreated control nucleoli (lane 2) were separated by SDS-PAGE and stained with Coomassie blue. Proteins that are more abundant in lane 1 were analyzed by MS and identified as labeled (numbers in brackets are SwissProt accession numbers). One of the identified proteins, p68, was tagged with YFP and expressed in HeLa cells. YFP-p68 (green signal) was not significantly detectable in nucleoli isolated from the YFP-p68 stable cell line (red signal is Pyronin Y staining) (E) but was found at caps following Actinomycin D treatment (F). Scale bars, 10 μm (A and B); 25 μm (C); and 15 μm (E and F).

reported so far for a single organelle and identifies 80 novel human genes that encode putative nucleolar proteins. Importantly, the proteomic analysis of nucleoli isolated from cells in which transcription had been inhibited revealed the dynamic nature of the nucleolar proteome. A subset of 11 nucleolar proteins was shown to increase their association with nucleoli following treatment of cells with Actinomycin D. This analysis provides new insights into the complexity of proteins in the nucleolus and identifies many nucleolar factors that are conserved in evolution from budding yeast to humans. The data indicate that the nucleolus is an organelle with a high degree of functional complexity.

The large-scale characterization and identification of nucleolar proteins required the application of a variety of parallel approaches for both the effective separation of different classes of proteins and their subsequent analysis by both MALDI TOF and nanoES MS. The identities of the proteins included in the nucleolar database were each confirmed through the detection of at least two independent peptides. Furthermore, in many cases, genes were identified from the MS data through direct online searching of the draft human genome sequence. The strategy used, involving real-time database searching during data acquisition, facilitated the identification of proteins in complex peptide mixtures. This provided major practical advantages in dealing with such a large-scale analysis as that involved in determining the nucleolar proteome. Previously, we demonstrated that 18 novel human proteins analyzed in a large-scale proteomic study of the spliceosome could be identified by screening EST databases [17]. Now, with access to the human genome sequence, we have demonstrated that searches directly in the human genome can be advantageous when identifying novel proteins from peptide mixtures. In these searches, different peptides were immediately linked by mapping to a confined region in the genome. In addition, the resulting information, which defines detailed gene structures, can play an important role in both extending and refining the annotation of the human genome. In this way, the predicted coding products of the genome can be compared with the actual primary structures of the cognate proteins expressed in vivo.

For over half of the known proteins identified in the purified nucleoli, evidence exists in the literature to support their nucleolar association. We are confident that the majority of the additional proteins identified, including those encoded by novel genes, are also bona fide nucleolar factors. Few known contaminants were detected in the nucleolar preparations, and among the novel proteins we observed low scores for signal peptide sequences that could be indicative of contaminating cytoplasmic and ER proteins. The surprisingly high proportion of novel proteins also argue for highly enriched preparations. Furthermore, transient transfection analysis of 18 selected genes from the list in Table S1, each tagged with YFP, showed that 15 associated with nucleoli, as judged by fluorescence microscopy. Extrapolating from this YFP-tagging analysis, we estimate that at least 80% of the unverified genes are likely to encode nucleolar proteins. In fact, this figure may underestimate the fraction of genes encoding genuine nucleolar proteins for several reasons. First, the presence of the YFP tag in the fusion protein may interfere with nucleolar localization. Second, some proteins are known only to localize to nucleoli at specific stages of the cell cycle [46, 31] or under specific metabolic conditions (this study). Third, because of the high sensitivity of protein detection by MS, our analysis may have identified proteins where

only a small fraction of the total protein is localized to nucleoli at one specific time and therefore may not be readily detected by fluorescence microscopy (for example, see [37]). Therefore, at least some of the proteins analyzed here that did not show obvious accumulation in nucleoli when fused to YFP may nonetheless correspond to endogenous nucleolar factors. Correspondingly, we anticipate that the great majority, if not all, of the proteins listed in Table S1 are nucleolar proteins.

The nucleolar proteome includes a wide range of different proteins (see Figure 5 and Table S1). The most common motifs found in these proteins are nucleic acid and nucleotide binding domains. The DEAD-box helicase motifs characteristic of the superfamily of RNA-dependent ATPases were present in 22 proteins, suggesting that the control of RNA base pairing interactions may be an important feature of nucleolar function. This conclusion is supported by the recent work of Bickmore and colleagues who found using a gene-trap method that the DEAD box motif was the most common motif found in the pool of nucleolar proteins detected [47]. A more detailed bioinformatic analysis of the nucleolar proteome is currently in progress and will be presented elsewhere. However, preliminary analysis indicates that there are no simple targeting motifs shared by all the nucleolar proteins, suggesting that their localization to the nucleolus involves a variety of mechanisms.

The major function known for the nucleolus is the transcription and processing of rRNAs and their subsequent assembly into ribosomal subunits. Consistent with this, the nucleolar proteome includes many ribosomal proteins, processing factors, and components required for transcription of the rRNA gene clusters, as well as homologs of genes involved in these processes in other organisms. This analysis has also identified novel proteins that may be involved in these processes, such as NNP8, a protein related to the yeast pre-rRNA processing factor Ski6p, and the putative exoribonucleases NNP15 and NNP11. However, not all the proteins appear to have functions associated with the known roles of the nucleolus. For example, we have also identified several protein translation factors, including eIF4A, eIF5A, eIF6, and ETF1 (peptide chain release factor subunit 1), which bind to active ribosomes in the cytoplasm but were not known to preassemble in the nucleolus with the individual ribosomal subunits [48]. Interestingly, eIF6 was recently reported to localize to nucleoli in mast cells [49], consistent with the observation reported here for HeLa cells. The presence of these translation factors raises the interesting possibility of nucleolar/nuclear translation activity [50, 51].

The high protein complexity of the nucleolus revealed in Table S1 implies that either the biogenesis of ribosomes is a surprisingly complex process and/or that the nucleolus carries out additional functions, consistent with the theory of a plurifunctional nucleolus [24, 26]. One of these additional roles may be in regulating the localization of nuclear factors in a cell cycle-dependent manner and thereby controlling their access to interaction partners [22, 31]. The purified nucleoli studied in this project were isolated from unsynchronized cells and therefore the proteins listed in Table S1 will include factors that may associate with nucleoli only at specific stages of the cell cycle. For example, among the proteins identified, the DEAD box protein p68 associates with nucleoli specifically at telophase [52], while the Bloom's Syndrome factor (BLM) specifically accumulates in nucleoli only during S phase [53]. It will be interesting in future to determine whether any of the novel protein factors show similar properties.

In addition to cell cycle-specific associations, other nuclear factors can show facultative interactions with the nucleolus that depend upon the metabolic state of the cell. This was illustrated in this study by the 11 proteins that associate with nucleoli after treatment of cells with the transcription inhibitor Actinomycin D. In the accompanying article by Fox et al. (this issue of *Current Biology*), we describe the detailed characterization of one such protein identified here, the novel factor PSP1 [37]. PSP1 shows a nucleolar relationship affected by the transcription state of the cell. Photobleaching analyses have revealed that PSP1 constantly traffics through the nucleolus yet has a steady-state accumulation within novel nucleoplasmic structures called "paraspeckles." However, when transcription is inhibited, PSP1 rapidly relocalizes to the nucleolar periphery where it accumulates within perinucleolar caps. This is consistent with the MS data reported here showing PSP1 is enriched in preparations of purified nucleoli from Actinomycin D-treated cells (Figure 6). These data are also in agreement with recent evidence showing other nuclear proteins are dynamic and can traffic rapidly between different nuclear compartments [8]. It is interesting that the 11 proteins found here enriched in the preparations of nucleoli purified from Actinomycin D-treated cells have obvious features in common. Many of the proteins have RNA binding motifs, and several have been implicated in similar modes of transcriptional control. For example, p72 and p68 can act as estrogen receptor coactivators [54, 55]; PSF [56] and p54nrb can act as corepressors interacting with the DNA binding domains of nuclear hormone receptors; and PSP2 (also called CoAA) coactivates transcription and interacts with a thyroid hormone receptor binding protein [57]. These data suggest a possible novel role for the nucleolus, involving the cycling of transcription factors, and it will be interesting to investigate this further in the future. It is likely that further analysis will also identify additional factors whose association with nucleoli is either increased or decreased by Actinomycin D treatment.

The proteins listed in Table S1 represent a core of major nucleolar proteins, most of which were detected multiple times in independent preparations of nucleoli. Clearly, it does not include every human protein that may associate or interact with nucleoli in vivo, and some previously reported nucleolar proteins were not detected. Weakly associated factors may be lost during the purification protocol, while other proteins may interact with nucleoli only under specific metabolic conditions that were not sampled in our study. Very low abundance proteins or factors with unusual structures or modifications may have escaped detection with the methods we have used so far. Our analysis of HeLa cell nucleoli will also exclude some cell type-specific nucleolar proteins. We intend to extend our coverage of nucleolar proteins by further proteomic analysis of nucleoli

purified from a variety of sources, including primary cells and cell lines derived from different tissues. In addition, it is anticipated that future improvements in the sensitivity of protein detection and analysis methods will enable the identification of additional nucleolar proteins.

In conclusion, although more work remains to be done, we believe that the nucleolar proteome detailed here already represents a significant advance toward defining a comprehensive inventory of nucleolar proteins. These data should be of value for future studies on the range of biological roles performed by the nucleolus as well as the mechanisms involved in its assembly and function.

Experimental Procedures

Nucleolar Isolation

Nucleoli were prepared from HeLa cell nuclei (Computer Cell Culture Centre, Belgium), using a method based on that first described by Muramatsu and coworkers in 1963 [58]. Aliquots (250 μl) containing ~1 × 10^8 nuclei were washed three times with PBS, resuspended in 5 ml buffer A (10 mM HEPES-KOH [pH 7.9], 1.5 mM $MgCl_2$, 10 mM KCL, 0.5 mM DTT), and dounce homogenized ten times using a tight pestle. Dounced nuclei were centrifuged at 228 × *g* for 5 min at 4°C. The nuclear pellet was resuspended in 3 ml 0.25 M sucrose, 10 mM $MgCl_2$, and layered over 3 ml 0.35 M sucrose, 0.5 mM $MgCl_2$, and centrifuged at 1430 × *g* for 5 min at 4°C. The clean, pelleted nuclei were resuspended in 3 ml 0.35 M sucrose, 0.5 mM $MgCl_2$, and sonicated for 6 × 10 s using a microtip probe and a Misonix XL 2020 sonicator at power setting 5. The sonicate was checked using phase contrast microscopy, ensuring that there were no intact cells and that the nucleoli were readily observed as dense, refractile bodies. The sonicated sample was then layered over 3 ml 0.88 M sucrose, 0.5 mM $MgCl_2$ and centrifuged at 2800 × *g* for 10 min at 4°C. The pellet contained the nucleoli, while the supernatant consisted of the nucleoplasmic fraction. The nucleoli were then washed by resuspension in 500 μl of 0.35 M sucrose, 0.5 mM $MgCl_2$, followed by centrifugation at 2000 × *g* for 2 min at 4°C.

Microscopy and Immunostaining of Samples

Whole HeLa cells were grown on coverslips, washed in PBS, then fixed for 5 min with paraformaldehyde (4% in 10 mM pipes [pH 6.8], 100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, and 2 mM EDTA) and immunolabeled. Permeabilization was performed with 1% Triton X-100 in PBS for 15 min at room temperature. Immunofluorescence labeling was carried out on fixed permeabilized cells by 1 hr incubation in 100 μl primary antibody (diluted in PBST [PBS plus 0.05% Tween 20]), followed by three washes in PBST and incubation for 1 hr in 100 μl secondary antibody (diluted in PBST) followed by three washes in PBST. Antibodies used were anti-p80 coilin monoclonal 5P10 (dilution 1:10) and anti-nucleolin 7G2 (1:1000, kindly provided by G. Dreyfuss) and anti-mouse FITC-conjugated secondary antibody (Jackson Lab, 1:250). Prior to mounting, samples were stained with 0.66 mM Pyronin Y (Sigma) for 2 s, then mounted using DABCO containing Mowiol. Purified nucleoli were immobilized on poly-L-lysine slides (BDH) and air dried. After rehydration with PBS (5 min), the nucleoli were incubated with monoclonal anti-nucleolin antibody 7G2 (1:100) or monoclonal anti-coilin antibody 5P10 (1:5) for 30 min. The slides were washed with PBS (3 × 5 min) and incubated with FITC-conjugated secondary antibody (Jackson Lab, 1:250) for 30 min. The nucleoli were counterstained with 0.66 mM Pyronin Y (Sigma) for 1 min. After washing as above, the nucleoli were embedded in DABCO containing Mowiol. Images were obtained using a Zeiss LSM 410 confocal laser scanning microscope.

Isolated nucleoli were processed for transmission electron microscopy using standard methods. In brief, nucleoli were centrifuged at 5000 rpm for 2 min. The pelleted nucleoli were washed briefly in PBS, fixed in 3.7% paraformaldehyde in PBS for 20 min at room temperature, washed three times in PBS, washed in water, postfixed in 1% osmium tetroxide in water for 15 min at room temperature, washed three times in water, dehydrated in 70% ethanol for 2 × 5 min, contrasted with 3% uranyl acetate in 70% ethanol for 15 min at room temperature, washed in 70% ethanol further dehydrated through 90% ethanol, 1 × 5 min, 100% ethanol, 3 × 5 min, propylene oxide, 2 × 5 min, 50:50 propylene oxide:epoxy resin mix for 2 hr at room temperature before embedding in epoxy resin. Sections were cut using a Reichart ultracut ultramicrotome and visualized in a JOEL 1200EX TEM.

Electrophoresis and Immunoblotting

For 1D SDS/PAGE, purified nucleoli were dissolved in 1× LDS sample buffer (Novex) + 100 mM DTT and heated at 70°C for 10 min. Nucleolar proteins were then separated on 3%–8% SDS Tris-Acetate gels and 4%–12% Bis-Tris gels (Novex) and stained with Coomassie colloidal blue according to the manufacturer's instructions (Novex). For 2D SDS/PAGE, IPG strips (Pharmacia) were in-gel rehydrated with samples dissolved in 2 M thiourea, 6 M urea, 1% CHAPS, 0.4% DTT, and 0.5% Pharmalyte (either 3-10 or 4-7). Proteins were focused for 40000 V in the first dimension, separated on vertical 12.5% SDS gels in the second dimension, and stained with silver. Protein slices/spots were excised, deposited in 96-well plates, in-gel reduced, alkylated with iodoacetamide, and digested with trypsin as previously described [59]. The resulting peptides were analyzed by MALDI-MS and nanoES tandem MS (see below).

For immunoblotting, protein samples separated by 1D SDS/PAGE (see above) were subsequently transferred onto nitrocellulose membrane using a submarine system (Novex) and buffer containing 12 mM Tris, 100 mM glycine, and 20% methanol. Following blocking with 5% milk powder in PBS + 0.05% Tween 20, the membranes were incubated with one of the following antibodies: rabbit anti-fibrillarin, Fib 42 (1:10000, gift from F. Fuller-Pace), mouse monoclonal anti-nucleolin, 7G2 (1:1000, G. Dreyfuss), mouse monoclonal anti-Lamin B, LN43.2, (gift from E.B. Lane), and nucleoporin NUP62 was detected using mouse monoclonal 414 (1:5000, Babco), then bound antibody was probed using anti-rabbit HRP conjugate (1:2000 dilution) or anti-mouse HRP conjugate (1:5000 dilution) (Pierce Chemical Co.) in PBS containing 5% milk powder and 0.05% Tween 20, and detected via chemiluminescence with ECL (Amersham Pharmacia Biotech).

Mass Spectrometry

MALDI mass spectra were acquired automatically on a Bruker REFLEX III reflectron time of flight (TOF) mass spectrometer (Bruker-Franzen, Bremen, Germany) as previously described [60, 61]. Peptides were also analyzed by nanoES tandem MS on a quadrupole time-of-flight mass spectrometer (QSTAR Pulsar, PE Sciex, Toronto, Canada) equipped with a nanoES ion source (MDS-Proteomics, Denmark). The peptide mixtures from in-gel digests were purified and concentrated prior to nanoESMS as described [62], except that the remaining supernatant were loaded on columns in parallel from 96-well plates.

Databases and Searching

Nonredundant (nrdb), predicted (IPI.1), expressed sequence tags and finished and unfinished human genome sequences (phases 0-3) were downloaded (ftp://ncbi.nlm.nih.gov/genbank and ftp://ftp.ensembl.org/IPI/) and converted into FASTA formatted sequence index files accepted by the PepSea database search software system (MDS proteomics, Denmark). Peptide masses measured by MALDI-MS were searched with 40 ppm mass accuracy, and "used" peptide masses were subtracted prior to repeated rounds of database searches. Peptide sequence tags [36] were assigned from tandem mass spectra assisted by the Inspector software (MDS Proteomics, Denmark) and used for database searching in real time against the respective databases using PepSea software. Calculated peptide ion and fragment ion masses of retrieved sequences were displayed in the mass spectra to establish unambiguous identification and to identify additional proteins by selection of unexplained peaks for MS/MS. Genome searching were performed as recently described [38].

Gene Prediction

Two WWW-based gene prediction programs were employed for further characterization of identified coding regions of the human

genome: GENSCAN, at the Massachusetts Institute of Technology (MIT, Boston, http://genes.mit.edu/GENSCAN.html), and HMMgene, at the Centre or Biological Sequence Analysis (CBS, The Technical University of Denmark, Lyngby, Denmark, http://www.cbs.dtu.dk/services/HMMgene). Nucleotide sequences corresponding to identified peptides were defined as coding and used as constraints in HMMgene predictions. The GPMAW software (Lighthouse Data, Denmark) were modified to calculate m/z values of exon-exon spanning peptides.

Cloning Nucleolar Proteins, Tagging with Fluorescent Protein, and Expression in HeLa Cells

Full-length cDNAs encoding selected nucleolar candidate proteins were amplified by PCR using specific ESTs (UK HGMP Resource Centre) or HeLa cDNA library (Clontech) as templates. Each construct was cloned into the pEYFP-C1 vector (Clontech), using restriction sites engineered onto the 5′ ends of each amplification primer. Details of ESTs, primer sequences, and restriction sites are available on request.

HeLa cells were grown in DMEM supplemented with 10% fetal calf serum and 1% penicillin streptomycin (Life Technologies). Fluorescent protein fusion constructs were transfected into HeLa cells seeded onto coverslips using Effectene reagent (Qiagen). At 16–18 hr posttransfection, the cells were washed in PBS and fixed in paraformaldehyde as above. Coverslips were mounted in Mowiol/Dabco onto slides for imaging. Images were obtained using a Zeiss LSM 410 confocal laser scanning microscope.

Actinomycin D Treatment of HeLa Cells

To study the changes in nucleolar composition caused by inhibition of transcription, we incubated HeLa cells in 5 μg/ml Actinomycin D (Sigma) diluted in DMEM (Gibco) with 10% fetal calf serum (Gibco). After 3 hr, the cells were harvested by trypsinization, washed three times with cold PBS, then resuspended at ~5 × 10^7/ml with 10 mM HEPES-KOH (pH 7.9), 10 mM KCl, 1.5 mM MgCl2, 0.5 mM DTT. The cell suspension was incubated on ice for 5 min and then disrupted by dounce homogenation (ten strokes) on ice. The nuclei released by homogenation were pelleted by centrifugation at 1000 × *g* for 5 min. Nucleoli were isolated from these nuclei as described above.

Supplementary Material

Supplementary Material including a table of proteins identified from nucleoli isolated from cultured human cells can be found at http://images.cellpress.com/supmat/supmatin.htm.

Acknowledgments

A.I.L. is a Wellcome Trust Principal Research Fellow and is funded by a Wellcome Trust Programme grant. The authors thank Peter Højrup for modifying GPMAW. Work in M.M.'s laboratory is funded by a Danish National Research Foundation grant to the Centre for Experimental BioInformatics. A.K.L.L. is funded by a Croucher studentship; Y.W.L. is funded by a Croucher postdoctoral fellowship; A.H.F. is funded by a Wellcome Trust International Travelling Fellowship; and C.E.L. is funded by the Wellcome Trust.

Received: October 15, 2001
Revised: November 14, 2001
Accepted: November 14, 2001
Published: January 8, 2002

References

1. Misteli, T., and Spector, D.L. (1998). The cellular organization of gene expression. Curr. Opin. Cell Biol. *10*, 323–331.
2. Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science *280*, 547–553.
3. Lewis, J.D., and Tollervey, D. (2000). Like attracts like: getting RNA processing together in the nucleus. Science *288*, 1385–1389.
4. Dundr, M., and Misteli, T. (2001). Functional architecture in the cell nucleus. Biochem. J. *356*, 297–310.
5. Matera, A.G. (1999). Nuclear bodies: multifaceted subdomains of the interchromatin space. Trends Cell Biol. *9*, 302–309.
6. Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. Nature *404*, 604–609.
7. Kruhlak, M.J., Lever, M.A., Fischle, W., Verdin, E., Bazett-Jones, D.P., and Hendzel, M.J. (2000). Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments. J. Cell Biol. *150*, 41–51.
8. Misteli, T. (2001). Protein dynamics: implications for nuclear architecture and gene expression. Science *291*, 843–847.
9. Platani, M., Goldberg, I., Swedlow, J.R., and Lamond, A.I. (2000). In vivo analysis of Cajal body movement, separation, and joining in live human cells. J. Cell Biol. *151*, 1561–1574.
10. Swedlow, J.R., and Lamond, A.I. (2001). Nuclear dynamics: where genes are and how they got there. Genome Biol. *2*. Published online March 9, 2001. reviews0002.1-0002.7.
11. Lamond, A., and Mann, M. (1997). Cell biology and the genome projects—a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. Trends Cell Biol. *7*, 139–142.
12. Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837–846.
13. Blackstock, W.P., and Weir, M.P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. *17*, 121–127.
14. Aebersold, R., and Goodlett, D.R. (2001). Mass spectrometry in Proteomics. Chem. Rev. *101*, 269–295.
15. Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S., and Spector, D.L. (1999). Purification and biochemical characterization of interchromatin granule clusters. EMBO J. *18*, 4308–4320.
16. Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., and Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. Proc. Natl. Acad. Sci. USA *94*, 385–390.
17. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nat. Genet. *20*, 46–50.
18. Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol. *148*, 635–651.
19. Shaw, P.J., and Jordan, E.G. (1995). The nucleolus. Annu. Rev. Cell Dev. Biol. *11*, 93–121.
20. Scheer, U., and Hock, R. (1999). Structure and function of the nucleolus. Curr. Opin. Cell Biol. *11*, 385–390.
21. Venema, J., and Tollervey, D. (1999). Ribosome synthesis in Saccharomyces cerevisiae. Annu. Rev. Genet. *33*, 261–311.
22. Carmo-Fonseca, M., Mendes-Soares, L., and Campos, I. (2000). To be or not to be in the nucleolus. Nat. Cell Biol. *2*, E107–112.
23. Dundr, M., Misteli, T., and Olson, M.O. (2000). The dynamics of postmitotic reassembly of the nucleolus. J. Cell Biol. *150*, 433–446.
24. Pederson, T. (1998). The plurifunctional nucleolus. Nucleic Acids Res. *26*, 3871–3876.
25. Olson, M.O., Dundr, M., and Szebeni, A. (2000). The nucleolus: an old factory with unexpected capabilities. Trends Cell Biol. *10*, 189–196.
26. Schneiter, R., Kadowaki, T., and Tartakoff, A.M. (1995). mRNA transport in yeast: time to reinvestigate the functions of the nucleolus. Mol. Biol. Cell *6*, 357–370.
27. Politz, J.C., Yarovoi, S., Kilroy, S.M., Gowda, K., Zwieb, C., and Pederson, T. (2000). Signal recognition particle components in the nucleolus. Proc. Natl. Acad. Sci. USA *97*, 55–60.
28. Lange, T.S., and Gerbi, S.A. (2000). Transient nucleolar localization of U6 small nuclear RNA in Xenopus Laevis oocytes. Mol. Biol. Cell *11*, 2419–2428.
29. Mitchell, J.R., Wood, E., and Collins, K. (1999). A telomerase component is defective in the human disease dyskeratosis congenita. Nature *402*, 551–555.
30. Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R.H., and Engelke, D.R. (1998). Nucleolar localization of early tRNA processing. Genes Dev. *12*, 2463–2468.

31. Visintin, R., and Amon, A. (2000). The nucleolus: the magician's hat for cell cycle tricks. Curr. Opin. Cell Biol. *12*, 752.
32. Huang, S. (2000). Review: perinucleolar structures. J. Struct. Biol. *129*, 233–240.
33. Gall, J.G. (2000). Cajal bodies: the first 100 years. Annu. Rev. Cell Dev. Biol. *16*, 273–300.
34. Vincent, W.S. (1955). Structure and chemistry of nucleoli. Int. Rev. Cytol. *4*, 269–298.
35. Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., and Mann, M. (1996). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. USA *93*, 14440–14445.
36. Mann, M., and Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal. Chem. *66*, 4390–4399.
37. Fox, A.H., Lam, Y.W., Leung, A., Lyon, C.E., Andersen, J.S., Mann, M., and Lamond, A.I. (2002). Paraspeckles: A novel nuclear domain. Curr. Biol. *12*, this issue, 13–25.
38. Kuster, B., Mortensen, P., Andersen, J.S., and Mann, M. (2001). Mass spectrometry allows direct identification of proteins in large genomes. Proteomics *5*, 641–650.
39. Zirwes, R.F., Eilbracht, J., Kneissel, S., and Schmidt-Zachmann, M.S. (2000). A novel helicase-type protein in the nucleolus: protein NOH61. Mol. Biol. Cell *11*, 1153–1167.
40. Strezoska, Z., Pestov, D.G., and Lau, L.F. (2000). Bop1 is a mouse WD40 repeat nucleolar protein involved in 28S and 5.8S RRNA processing and 60S ribosome biogenesis. Mol. Cell. Biol. *20*, 5516–5528.
41. Tanner, N.K., and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol. Cell *8*, 251–262.
42. Ghosh, S. (1976). The nucleolar structure. Int. Rev. Cytol. *44*, 1–28.
43. Raska, I., Ochs, R.L., Andrade, L.E., Chan, E.K., Burlingame, R., Peebles, C., Gruol, D., and Tan, E.M. (1990). Association between the nucleolus and the coiled body. J. Struct. Biol. *104*, 120–127.
44. Carmo-Fonseca, M., Pepperkok, R., Carvalho, M.T., and Lamond, A.I. (1992). Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. J. Cell Biol. *117*, 1–14.
45. Dye, B.T., and Patton, J.G. (2001). An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles. Exp. Cell Res. *263*, 131–144.
46. Xirodimas, D., Saville, M.K., Edling, C., Lane, D.P., and Lain, S. (2001). Different effects of p14ARF on the levels of ubiquitinated p53 and Mdm2 in vivo. Oncogene *20*, 4972–4983.
47. Sutherland, H.G., Mumford, G.K., Newton, K., Ford, L.V., Farrall, R., Dellaire, G., Caceres, J.F., and Bickmore, W.A. (2001). Large-scale identification of mammalian proteins localized to nuclear sub-compartments. Hum. Mol. Genet. *10*, 1995–2011.
48. Pestova, T.V., Kolupaeva, V.G., Lomakin, I.B., Pilipenko, E.V., Shatsky, I.N., Agol, V.I., and Hellen, C.U. (2001). Molecular mechanisms of translation initiation in eukaryotes. Proc. Natl. Acad. Sci. USA *98*, 7029–7036.
49. Oh, C.K., Filler, S.G., and Cho, S.H. (2001). Eukaryotic translation initiation factor-6 enhances histamine and IL-2 production in mast cells. J. Immunol. *166*, 3606–3611.
50. Pederson, T., and Politz, J.C. (2000). The nucleolus and the four ribonucleoproteins of translation. J. Cell Biol. *148*, 1091–1095.
51. Iborra, F.J., Jackson, D.A., and Cook, P.R. (2001). Coupled transcription and translation within nuclei of mammalian cells. Science *293*, 1139–1142.
52. Nicol, S.M., Causevic, M., Prescott, A.R., and Fuller-Pace, F.V. (2000). The nuclear DEAD box RNA helicase p68 interacts with the nucleolar protein fibrillarin and colocalizes specifically in nascent nucleoli during telophase. Exp. Cell Res. *257*, 272–280.
53. Sanz, M.M., Proytcheva, M., Ellis, N.A., Holloman, W.K., and German, J. (2000). BLM, the Bloom's syndrome protein, varies during the cell cycle in its amount, distribution, and co-localization with other nuclear proteins. Cytogenet. Cell Genet. *91*, 217–223.
54. Endoh, H., Maruyama, K., Masuhiro, Y., Kobayashi, Y., Goto, M., Tai, H., Yanagisawa, J., Metzger, D., Hashimoto, S., and Kato, S. (1999). Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha. Mol. Cell. Biol. *19*, 5363–5372.
55. Watanabe, M., Yanagisawa, J., Kitagawa, H., Takeyama, K., Ogawa, S., Arao, Y., Suzawa, M., Kobayashi, Y., Yano, T., Yoshikawa, H., et al. (2001). A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. EMBO J. *20*, 1341–1352.
56. Mathur, M., Tucker, P.W., and Samuels, H.H. (2001). PSF is a novel corepressor that mediates its effect through Sin3A and the DNA binding domain of nuclear hormone receptors. Mol. Cell. Biol. *21*, 2298–2311.
57. Iwasaki, T., Chin, W.W., and Ko, L. (2001). Identification and characterization of rrm-containing coactivator activator (coaa) as trbp-interacting protein, and its splice variant as a coactivator modulator (coam). J. Biol. Chem. *276*, 33375–33383.
58. Muramatsu, M., Smetana, K., and Busch, H. (1963). Quantitative aspects of isolation of nucleoli of the Walker carcinosarcoma and liver of the rat. Cancer Res. *25*, 693–697.
59. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. *68*, 850–858.
60. Vorm, O., Roepstorff, P., and Mann, M. (1994). Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. Anal. Chem. *66*, 3281–3287.
61. Jensen, O.N., Mortensen, P., Vorm, O., and Mann, M. (1997). Automation of matrix-assisted laser desorption/ionization mass spectrometry using fuzzy logic feedback control. Anal. Chem. *69*, 1706–1714.
62. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano- electrospray mass spectrometry. Nature *379*, 466–469.

Supplementary Material

Directed Proteomic Analysis of the Human Nucleolus

Jens S. Andersen, Carol E. Lyon, Archa H. Fox, Anthony K.L. Leung, Yun Wah Lam, Hanno Steen, Matthias Mann, and Angus I. Lamond

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
NNP1	Novel Nucleolar Protein	9459	Hs.63908	84324						
H4F2	H4 histone, family 2	11311	Hs.123053	3022	Nuclear protein; Chromosomal protein; Nucleosome core; DNA-binding;Multigene family.					
H2BFA	H2B histone family, member A	13775	Hs.247817	8339	Nuclear protein; Chromosomal protein; Nucleosome core; DNA-binding;Multigene family.	*	HTB1			
SNRPD3	small nuclear ribonucleoprotein D3 polypeptide (18kD)	13916	Hs.1575	6634	Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;Systemic lupus erythematosus; Repeat.	*	SMD3			
H2AFE	H2A histone family, member E	13936	Hs.274590	8331	Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding.		HTA2			
NHP2L1/NHPX	non-histone chromosome protein 2 (S. cerevisiae)-like 1	14070	Hs.182255	4809		1,2*	SNU13			
RPL31	ribosomal protein L31	14462	Hs.184014	6160	Ribosomal protein.	r	RPL31A			
RPL22	ribosomal protein L22	14655	Hs.326249	6146	Ribosomal protein; RNA-binding; Heparin-binding.	r	RPL22A			
SRP14	signal recognition particle 14kD (homologous Alu RNA-binding protein)	14658	Hs.180394	6727						
RPS12	Ribosomal protein S12	14737	Hs.339696	6206		r	RPS12			
NNP2	Novel Nucleolar Protein	15433	ctg17711_56							
NNP3	Novel Nucleolar Protein	15524	Hs.38114	55299	putative ribosome biogenesis (based on similarity with yeast protein)					
RPL27	ribosomal protein L27	15666	Hs.111611	6155	Ribosomal protein.	r*	RPL27B			
RPS16	ribosomal protein S16	16314	Hs.80617	6217	Ribosomal protein.	r	RPS16A			
RPL27A	ribosomal protein L27a	16561	Hs.76064	6157	Ribosomal protein.	r	RPL28			
EIF5A	eukaryotic translation initiation factor 5A	16929	Hs.119140	1984	translation		HYP2			
RPS13	ribosomal protein S13	17091	Hs.165590	6207	Ribosomal protein.	r	RPS13			
POLR2H	polymerase (RNA) II (DNA directed) polypeptide H	17157	Hs.3128	5437	Transferase; DNA-directed RNA polymerase; Transcription;Nuclear protein.					
NOLA2	nucleolar protein family A, member 2	17201	Hs.23990	55651		2	NHP2			
RPL26	ribosomal protein L26	17258	Hs.91379	6154	Ribosomal protein.	r	RPL26B			

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
RPL23A	ribosomal protein L23a	17695	Hs.184776	6147	Ribosomal protein.	r	RPL25			
RPS18	ribosomal protein S18	17718	Hs.275865	6222	Ribosomal protein.	r	RPS18B			
RPL12	ribosomal protein L12	17989	Hs.182979	6136	Ribosomal protein; RNA-binding.	r	RPL12B			
PPIA	peptidylprolyl isomerase A (cyclophilin A)	18109	Hs.182937	5478			CPH1			
RPS11	ribosomal protein S11	18430	Hs.182740	6205	Ribosomal protein.	r	RPS11A			
NPM3	nucleophosmin/nucleoplasmin 3	19343	Hs.90691	10360						
SFRS3	splicing factor, arginine/serine-rich 3	19557	Hs.167460	6428		2				
TPT1	tumor protein, translationally-controlled 1	19709	Hs.279860	7178		+	YKL056C			
RPL11	ribosomal protein L11	20349	Hs.179943	6135	Ribosomal protein; rRNA-binding.	r	RPL11B			
ARPC3	actin related protein 2/3 complex, subunit 3 (21 kD)	20546	Hs.6895	10094			ARC18			
HP1g	heterochromatin-like protein 1	20839	Hs.278554	51563	Nuclear protein; Phosphorylation; 3D-structure.	*				
NNP4	Novel Nucleolar Protein	21148	Hs.279918	51491						
CSRP1	cysteine and glycine-rich protein 1	21292	Hs.108080	1465						
RPL17	ribosomal protein L17	21393	Hs.82202	6139	Ribosomal protein.	r	RPL17A			
CBX1	chromobox homolog 1 (Drosophila HP1 beta)	21415	Hs.77254	10951	Nuclear protein; Phosphorylation; 3D-structure.					
RPL18	ribosomal protein L18	21617	Hs.75458	6141	Ribosomal protein.	r	RPL18B			
MRPS4	hypothetical protein FLJ10968	21964	Hs.6118	55272		2	IMP3			
RPL9	ribosomal protein L9	21977	Hs.157850	6133	Ribosomal protein.	r	RPL9B			
PPIF	peptidylprolyl isomerase F (cyclophilin F)	22040	Hs.173125	10105			CPH1			
TASR1	TLS-associated serine-arginine protein 1	22336	Hs.288038	10772						

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
RPS9	ribosomal protein S9	22497	Hs.180920	6203	Ribosomal protein.	r	RPS9B			
NOLA1	nucleolar protein family A, member 1 (H/ACA small nucleolar RNPs)	22519	Hs.69851	54433		2	GAR1			
PPIB	peptidylprolyl isomerase B (cyclophilin B)	22741	Hs.699	5479	Cyclosporin; Isomerase; Rotamase; Signal; Endoplasmic reticulum;Multigene family; 3D-structure.		CYP5			
RPS5	ribosomal protein S5	22948	Hs.76194	6193	Ribosomal protein.	r	RPS5			
RPL13A	ribosomal protein L13a	23445	Hs.119122	23521	Ribosomal protein.	r	RPL16A			
RPL14	ribosomal protein L14	23803	Hs.738	9045	Ribosomal protein; Repeat.	r	RPL14B			
RPL15	ribosomal protein L15	24068	Hs.74267	6138	Ribosomal protein.	r	RPL15B			
RPL13	ribosomal protein L13	24348	Hs.180842	6137	Ribosomal protein.	r	RPL13B			
ALY	Transcriptional coactivator	24485	Hs.241520	10189						
RPS8	ribosomal protein S8	24490	Hs.151604	6202	Ribosomal protein.	r	RPS8B			
POLR2E	polymerase (RNA) II (DNA directed) polypeptide E (25kD)	24612	Hs.24301	5434	Transferase; DNA-directed RNA polymerase; Transcription;Nuclear protein.		RPB5			
NNP5	Novel Nucleolar Protein	24663	7228293							
RPL10A	ribosomal protein L10a	25030	Hs.252574	4736	Ribosomal protein.	r	RPL1B			
SFRS2	splicing factor, arginine/serine-rich 2	25575	Hs.73965	6427						
SFRS9	splicing factor, arginine/serine-rich 9	25656	Hs.77608	8683		2				
UBB	ubiquitin B	25818	Hs.183842	7314			UBI4			
NNP6	Novel Nucleolar Protein	25981	ctg18647_6							
NNP7	Novel Nucleolar Protein	26490	Hs.12045	10436			EMG1			
ITGB4BP/EIF6	integrin beta 4 binding protein	26590	Hs.5215	3692	Initiation factor; Protein biosynthesis.	1	CDC95			
NNP8	Novel Nucleolar Protein	26668	Hs.97574	54512		2	SKI6			

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
SFRS7	splicing factor, arginine/serine-rich 7 (35kD)	27594	Hs.184167	6432						
RPS0	60S acidic ribosomal protein PO	27619	Hs.274201	51154	weakly similar to 60S acidic ribosomal protein P0	r	MRT4			
RARG-1	retinoic acid repressible protein	27690	Hs.106346	51406						
SFRS1	splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)	27727	Hs.73737	6426		2				
RPL8	ribosomal protein L8	28253	Hs.178551	6132	Ribosomal protein.	r	RPL2B			
SNRPA1	small nuclear ribonucleoprotein polypeptide A'	28557	Hs.80506	6627						
NNP9	Novel Nucleolar Protein	28647	6002306							
RPS6	ribosomal protein S6	28748	Hs.241507	6194	Ribosomal protein.	r*	RPS6B			
NNP10	Novel Nucleolar Protein	29205	7023477				RRP4			
NNP11	Opa-interacting protein 2	29280	Hs.274170	11340		2	RRP45			
RPL7	ribosomal protein L7	29282	Hs.153	6129	Ribosomal protein; Repeat; RNA-binding.	r	RPL7A			
RPS4X	ribosomal protein S4, X-linked	29694	Hs.108124	6191	Ribosomal protein.	r*	RPS4A			
PHB	prohibitin	29861	Hs.75323	5245			PHB1			
RPS3A	ribosomal protein S3A	29888	Hs.77039	6189	Ribosomal protein.	r	RPS1B			
RRP40	CGI-102 protein	30028	Hs.177677	51010		2	YOL142w			
RPL7A	ribosomal protein L7a	30035	Hs.99858	6130	Ribosomal protein.	r	RPL8B			
NNP12	Novel Nucleolar Protein	30236	3881975				NSA2			
NNP13	Novel Nucleolar Protein	30504	Hs.111449	51118			YKL099c			
DNAJC8	splicing factor similar to dnaJ	30987	Hs.74711	22826						
ABT1	TATA-binding protein-binding protein	31136	Hs.109428	29777		2	YNR054C			

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
SFRS5	splicing factor, arginine/serine-rich 5	31378	Hs.166975	6430						
NNP14	Novel Nucleolar Protein	31586	ctg14521_16							
HNRPC	heterogeneous nuclear ribonucleoprotein C (C1/C2)	31966	Hs.182447	3183	Nuclear protein; RNA-binding; Ribonucleoprotein; Phosphorylation;Alternative splicing.					
NNP15	Novel Nucleolar Protein	32332	Hs.182877	23016						
NNP16	Novel Nucleolar Protein	32505	Hs.239934	27341						
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	32746	Hs.9614	4869		1*				
RPL6	ribosomal protein L6	32785	Hs.174131	6128	Ribosomal protein.	r	RPL6B			
LAMR1	laminin receptor 1 (67kD, ribosomal protein SA)	32854	Hs.181357	3921	Ribosomal protein; Repeat.	r	RPS0A			
RRP40	homolog of Yeast RRP4 (ribosomal RNA processing 4), 3'-5'-exoribonuclease	33017	Hs.211973	23404		2	RRP4			
HRB2	HIV-1 rev binding protein 2	33283	Hs.154762	11103		2	KRR1			
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isoform	33774	Hs.79081	5501		*	GLC7			
FBL	fibrillarin	33932	Hs.99853	2091	Nuclear protein; Methylation; Ribonucleoprotein; rRNA processing;RNA-binding; Antigen.	1,2*	NOP1			
HNRPA1	heterogeneous nuclear ribonucleoprotein A1	34196	Hs.249495	3178	Nuclear protein; RNA-binding; Repeat; Ribonucleoprotein;Methylation; Alternative splicing; 3D-structure.		HRP1			
RPL5	ribosomal protein L5	34544	Hs.180946	6125	Ribosomal protein.	r*	RPL5			
RPLP0	Ribosomal protein, large, P0	34588	Hs.194676	6175		r	RPP0			
NNP17	homolog of yeast EBNA1-binding protein	34820	Hs.74407	10969		2	EBP2			
BYSL	bystin-like	35171	Hs.106880	705			ENP1			
NNP18	Novel Nucleolar Protein	35711	Hs.9043	25983						
HNRPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	36006	Hs.232400	3181	Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;Alternative splicing.		HRP1			
HNRPAB	heterogeneous nuclear ribonucleoprotein A/B	36347	Hs.81361	3182			HRP1			

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
HNRPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)	36983	Hs.279681	3189						
RP1	retinitis pigmentosa 1 (autosomal dominant)	37031	Hs.251687	6101						
NNP19	Novel Nucleolar Protein	37035	Hs.99829	54984						
PPP1CB	protein phosphatase 1, catalytic subunit, beta isoform	37187	Hs.21537	5500	Hydrolase; Glycogen metabolism; Multigene family; Cell division.		GLC7			
HNRPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	37429	Hs.232400	3181			HRP1			
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isoform	37512	Hs.183994	5499	Hydrolase; Iron; Manganese; Glycogen metabolism; Multigene family;Cell division; Alternative splicing; 3D-structure.		GLC7			
PCBP1	poly(rC)-binding protein 1	38039	Hs.2853	5093			PBP2			
NNP20	Novel Nucleolar Protein	38397	Hs.59425	79707						
RPA40	RNA polymerase I subunit	38647	Hs.5409	9533		1*	RPC40			
HNRPA1	heterogeneous nuclear ribonucleoprotein A1	38828	Hs.249495	3178			HRP1			
ANXA2	annexin A2	38832	Hs.217493	302	Annexin; Calcium/phospholipid-binding; Repeat; Phosphorylation;Acetylation.					
NNP21	Novel Nucleolar Protein	38860	Hs.126522							
PMSCL1	polymyositis/scleroderma autoantigen 1 (75kD)	39235	Hs.91728	5393	Antigen; Nuclear protein.	1,2	RRP45			
H2AFY	H2A histone family, member Y	39489	Hs.75258	9555			HTA2			
RFC2	replication factor C (activator 1) 2 (40kD)	39613	Hs.139226	5982	DNA replication; ATP-binding; Nuclear protein.		RFC4			
NNP22	Novel Nucleolar Protein	40195	Hs.287863	80135		2	RPF1			
RBM4	RNA binding motif protein 4	40940	Hs.6106	5936	MHC.					
RNAC	RNA cyclase homolog	41327	Hs.113052	10171		2	RCL1			
ACTB	actin, beta	41736	Hs.288061	60			ACT1			
PSP1	Paraspeckles protein 1	41739	Hs.16364	55269		*				

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
RBMX/ HNRPG	RNA binding motif protein, X chromosome hnRNP G)	42240	Hs.146381	27316	Nuclear protein; RNA-binding; Ribonucleoprotein; Glycoprotein.					
NNP23	Novel Nucleolar Protein	42856	Hs.3487	54663						
NNP24/BING4	BING4	43479	Hs.17930	9277		*	YER082c			
PA2G4	proliferation-associated 2G4, 38kD	43787	Hs.5181	5036						
NNP25	Novel Nucleolar Protein	44090	Hs.71040	55646						
NNP26	KIAA0112 protein; homolog of yeast ribosome biogenesis regulatory protein RRS1	44502	Hs.71827	23212		2	RRS1			
ILF2	interleukin enhancer binding factor 2, 45kD	44697	Hs.75117	3608						
ACTR2	ARP2 (actin-related protein 2, yeast) homolog	44761	Hs.42915	10097	Structural protein; Cytoskeleton.		ARP2			
RPL3	Ribosomal protein L3	46109	Hs.119598	6122		r	RPL3			
NNP27	Novel Nucleolar Protein	46147	Hs.183253	65083						
EIF4A1	eukaryotic translation initiation factor 4A, isoform 1	46154	Hs.129673	1973	Initiation factor; Protein biosynthesis; ATP-binding; RNA-binding;DNA binding; Helicase.		TIF2			
P5	protein disulfide isomerase-related protein	46199	Hs.182429	10130	Isomerase; Redox-active center.					
IF4N	KIAA0111 gene product	46833	Hs.79768	9775	Initiation factor; Protein biosynthesis; ATP-binding; RNA-binding;DNA binding; Helicase.		FAL1			
NNP28	Novel Nucleolar Protein	46900	10438026				REX4			
NNP29	Novel Nucleolar Protein	47306	Hs.325321	57418						
ACTR3	ARP3 (actin-related protein 3, yeast) homolog	47371	Hs.5321	10096	Structural protein; Cytoskeleton.		ARP3			
NNP30	Novel Nucleolar Protein	47694	Hs.73291	55759			YTM1			
RPL4	ribosomal protein L4	47987	Hs.286	6124	Ribosomal protein.	r	RPL4B			
GRSF1	G-rich RNA sequence binding factor 1	47999	Hs.309763	2926	RNA-binding; Repeat.					
CSTF1	cleavage stimulation factor, 3' pre-RNA, subunit 1, 50kD	48357	Hs.172865	1477	Repeat; WD repeat; Nuclear protein.					

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
NNP31	Novel Nucleolar Protein	48564	Hs.17428	51742						
BAT1	HLA-B associated transcript-1	48991	Hs.55296	7919	ATP-binding; RNA-binding; Helicase; Nuclear protein.		SUB2			
ETF1	eukaryotic translation termination factor 1	49031	Hs.77324	2107	Protein biosynthesis.		SUP45			
HNRPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	49229	Hs.245710	3187	Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.					
NNP32	Novel Nucleolar Protein	49398	Hs.8768	55752			CDC3			
DDXL	nuclear RNA helicase, DECD variant of DEAD box family	49648	Hs.179606	10212	Helicase.		SUB2			
TUBB1	Homo sapiens TUBB1 gene for human beta tubulin 1, class VI	49759	Hs.303023	81027			TUB2			
TUFM	Tu translation elongation factor, mitochondrial	49884	Hs.12084	7284			TUF1			
TUBA6	tubulin, alpha 6	50151	14389308	84790			TUB1			
NNP33	Novel Nucleolar Protein	50932	12052855				RRP3			
HNRPK	heterogeneous nuclear ribonucleoprotein K	50976	Hs.129548	3190	Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;Alternative splicing; DNA-binding; Phosphorylation.					
TUBG1	tubulin, gamma 1	51198	Hs.21635	7283	Microtubules; GTP-binding.		TUB4			
NNP34	Novel Nucleolar Protein	51388	Hs.273344	25879			SOF1			
NNP35	Novel Nucleolar Protein	51479	Hs.5158	23378			YDR083w			
U3-55K	U3 snoRNP-associated 55-kDa protein	52468	Hs.153768	9136		1,2	RRP9			
NOP52	DNA segment on chromosome 21 (unique) 2056 expressed sequence	53067	Hs.110757	8568	Nuclear protein.	2	RRP1			
NNP36	Novel Nucleolar Protein	53471	ctg16922_16							
NNP37	peter pan (Drosophila) homolog	53479	Hs.65302	56342		*	SSF1			
KRT8	keratin 8	53562	Hs.242463	3856	Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin;Phosphorylation.					
U2AF65	U2 small nuclear ribonucleoprotein auxiliary factor (65kD)	53843	Hs.7655	11338	Nuclear protein; RNA-binding; mRNA splicing; Repeat.					

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
NONO/p54nrb	non-POU-domain-containing, octamer-binding	54231	Hs.172207	4841	RNA-binding; Nuclear protein; Repeat.	1				
NNP38	glioma tumor suppressor candidate region gene 2	54417	Hs.2237	29997		2*	YPL146C			
NNP39	disrupter of silencing 10	54672	Hs.322901	57050		*				
NNP40	Novel Nucleolar Protein	54854	Hs.143187	54555			DBP8			
NMP200	nuclear matrix protein NMP200 related to splicing factor PRP19	55181	Hs.173980	27339	Matrix protein.	+	TAF90			
NNP41	Novel Nucleolar Protein	55213	Hs.99423	11056	Helicase.		ROK1			
NNP42	nuclear phosphoprotein similar to S. cerevisiae PWP1	55828	Hs.172589	11137	Nuclear protein; Phosphorylation; Repeat; WD repeat.	*	PWP1			
ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	56560	Hs.25	506	ATP synthesis; CF(1); Hydrogen ion transport;Hydrolase; ATP-binding; Mitochondrion; Transit peptide.		ATP2			
GRP58	glucose regulated protein, 58kD	56782	Hs.289101	2923			PDI1			
NNP43/SAZD	Novel Nucleolar Protein	57074	Hs.114416	10607	Repeat; WD repeat.	*	YLR222C			
P4HB	protein disulfide isomerase	57116	Hs.75655	5034	Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal;3D-structure.		PDI1			
nPTB	neural polypyrimidine tract binding protein	57221	Hs.34956	58155		1				
NNP44	Novel Nucleolar Protein	57707	Hs.6153	51096			YJL069c			
DKC1	dyskeratosis congenita 1, dyskerin	58130	Hs.4747	1736	Nuclear protein.	1,2*	CBF5			
NNP45	Novel Nucleolar Protein	58131	Hs.2471	9933	Repeat; RNA-binding.		YDR496c			
KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	58204	Hs.159557	3838	Transport; Protein transport; Repeat; Nuclear protein.	2	SRP1			
NNP46	Novel Nucleolar Protein	58343	Hs.85963	26156		+				
NOP5/NOP58	nucleolar protein NOP5/NOP58	59578	Hs.119908	51602		2	NOP5			
CCT3	chaperonin containing TCP1, subunit 3 (gamma)	60901	Hs.1708	7203	Chaperone; ATP-binding; Multigene family.		CCT3			
HSPD1	heat shock 60kD protein 1 (chaperonin)	61024	Hs.79037	3329	Chaperone; ATP-binding; Mitochondrion; Transit peptide.		HSP60			

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
NOH61	putative nucleolar RNA helicase	61589	Hs.10098	54606		1*	DBP9			
COILIN	Coilin	62608	Hs.966	8161		1*				
DED	apoptosis antagonizing transcription factor	63133	Hs.16178	26574			BFR2			
NNP47	Novel Nucleolar Protein	64139	Hs.279923	26354			YNR053C			
NNP48	Novel Nucleolar Protein	64631	Hs.65234	55661	Helicase.	2	DRS1			
LMNA	lamin A/C	65135	Hs.77886	4000	Intermediate filament; Heptad repeat pattern; Coiled coil;Nuclear protein; Lipoprotein; Prenylation; Phosphorylation;Alternative splicing.					
LMNB1	lamin B1	66408	Hs.89497	4001	Intermediate filament; Heptad repeat pattern; Coiled coil;Nuclear protein; Lipoprotein; Prenylation; Phosphorylation.					
NOP56	nucleolar protein (KKE/D repeat)	67248	Hs.296585	10528	Nuclear protein.	1,2	SIK1 NOP56			
PES1	pescadillo (zebrafish) homolog 1, containing BRCT domain	68594	Hs.13501	23481		1*	YGR103W			
RBM14/PSP2	RNA binding motif protein 14	69491	Hs.11170	10432		2				
DDX5/P68	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68kD)	69661	Hs.76053	1655	ATP-binding; RNA-binding; Helicase; Nuclear protein.		DBP2			
HSPA1A	heat shock 70kD protein 1A	69868	Hs.8997	3303	ATP-binding; Chaperone; Heat shock; Multigene family; MHC III;3D-structure.		SSA2			
G22P1	thyroid autoantigen 70kD (Ku antigen)	69997	Hs.197345	2547	Helicase; Nuclear protein; DNA-binding; Phosphorylation; Antigen;Systemic lupus erythematosus; Acetylation.					
HSPA2	heat shock 70kD protein 2	70280	Hs.75452	3306	Heat shock.		SSA3			
HSPA8	heat shock 70kD protein 8	70898	Hs.180414	3312	ATP-binding; Heat shock; Multigene family.		SSA2			
HSPA5	heat shock 70kD protein 5 (glucose-regulated protein, 78kD)	70931	Hs.75410	3309			KAR2			
DDX17/P72	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 17 (72kD)	72371	Hs.343411	10521		1	DBP2			
NNP49	Novel Nucleolar Protein	72917	ctg595_12							
DDX3	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3	73243	Hs.147916	1654	Helicase; ATP-binding; RNA-binding; DNA-binding.		DED1			
NNP50	Novel Nucleolar Protein	73299	Hs.109045	55153			SDA1			

Continued

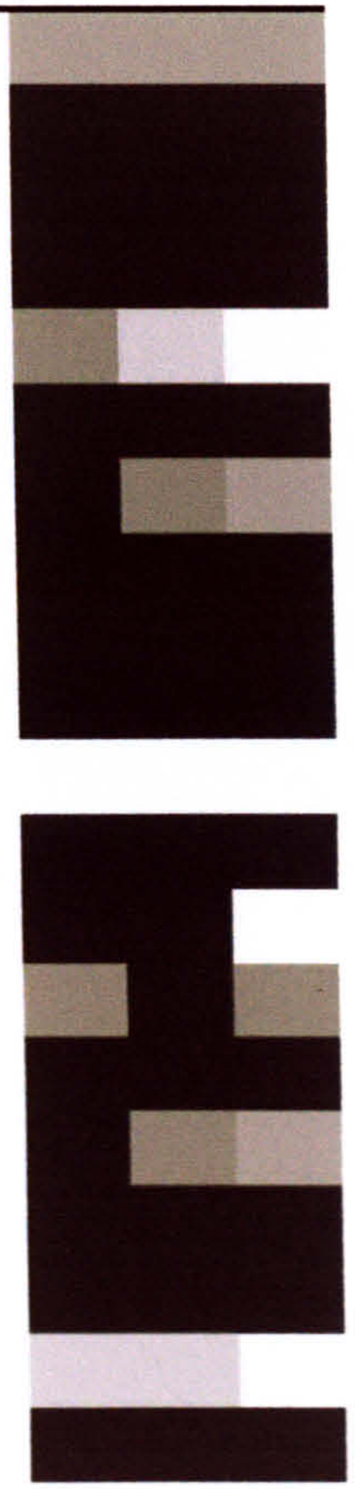
Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
NOLC1	nucleolar phosphoprotein p130	73720	Hs.75337	9221		1,2*	SRP40			
HSPA9B	heat shock 70kD protein 9B (mortalin-2)	73780	Hs.3069	3313	ATP-binding; Mitochondrion; Transit peptide.		SSC1			
PARN	poly(A)-specific ribonuclease (deadenylation nuclease)	73964	Hs.43445	5073						
NNP51	Novel Nucleolar Protein	74233	Hs.215766	23560		2*	NOG1			
NNP52	Novel Nucleolar Protein	75073	Hs.235498	79954			YGR145W			
DDX18	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated)	75753	Hs.100555	8886	Helicase; ATP-binding; RNA-binding.	2*	HAS1			
SFPQ/PSF	splicing factor proline/glutamine rich (polypyrimidine tract-binding protein-associated)	76149	Hs.180610	6421		1				
NCL	nucleolin	76270	Hs.79110	4691	Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat:RNA-binding.	1	PAB1			
NNP53	Novel Nucleolar Protein	76747	Hs.90315	23160			YDR398			
NNP54/BOP1	Novel Nucleolar Protein	77577	Hs.30736	23246	Repeat; WD repeat.	2	ERB1			
MPHOPSH10	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein	77682	Hs.201676	10199		1	MPP10			
HNRPM	heterogeneous nuclear ribonucleoprotein M	77755	Hs.79024	4670	Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;Alternative splicing.					
NNP55	Novel Nucleolar Protein	82093	10435709							
NNP56	Novel Nucleolar Protein	82502	Hs.152629	23076						
GU2	nucleolar protein GU2	82565	Hs.7392	79009		1	DBP3			
NNP57/NGB	Novel Nucleolar Protein	83655	Hs.75528	29889		1	YNR053C			
NNP58	Novel Nucleolar Protein	84878	Hs.134200	26155		2	YOR206W			
NNP59	Novel Nucleolar Protein	85766	Hs.274263	55131		2	NOP4			
NNP60	novel putative protein similar to YIL091C yeast hypothetical 84 kD protein from SGA1-KTR7	87397	Hs.194754	27042			YIL091C			
UBTF	transcription factor UBF	89406	Hs.89781	7343		1				

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
DDX21	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21	89764	Hs.169531	9188		1	DBP1			
TOP1	topoisomerase (DNA) I	91514	Hs.317	7150	Isomerase; Topoisomerase; DNA-binding; Polymorphism; 3D-structure.	1	TOP1			
DDX15	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 15	92829	Hs.5683	1665	mRNA processing; mRNA splicing; Helicase; ATP-binding;Nuclear protein.		PRP43			
NOL1	nucleolar protein 1 (120kD)	96213	Hs.15243	4839	Nuclear protein; Antigen.	1,2	NOP2			
NNP61	Novel Nucleolar Protein	99509	Hs.170114	23044		1				
DDX10	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 10 (RNA helicase)	101158	Hs.41706	1662	Helicase; ATP-binding; RNA-binding.	2*	HCA4			
PMSCL2	polymyositis/scleroderma autoantigen 2 (100kD)	101918	Hs.75584	5394	Hydrolase; Nuclease; Antigen; Nuclear protein; Alternative splicing.	1,2	RRP6			
NNP62	PWP2 (periodic tryptophan protein, yeast) homolog	103709	Hs.79380	5822	Repeat; WD repeat.	*	PWP2			
NNP63	Novel Nucleolar Protein	107183	Hs.33085	10885			DIP2			
NNP64	Novel Nucleolar Protein	107366	Hs.278496	9904			YPR112c			
NNP65	Novel Nucleolar Protein	110510	Hs.274149	27043	Nuclear protein.					
NNP66	Novel Nucleolar Protein	113913	14763765				YNL132w			
ADPRT	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)	114135	Hs.177766	142	Transferase; Glycosyltransferase; NAD; DNA-binding; Nuclear protein;ADP-ribosylation; Zinc-finger; Zinc.					
CBF2	CCAAT-box-binding transcription factor	114699	Hs.184760	10153	Transcription regulation; Activator; Nuclear protein.		YDR060W			
NNP67	Novel Nucleolar Protein	119871	Hs.278608	23517	Helicase; ATP-binding; Nuclear protein.	2	MTR4			
NNP68	Novel Nucleolar Protein	136384	Hs.60103	23223			YPL012W			
RAD50	RAD50 (S. cerevisiae) homolog	138432	Hs.41587	10111			RAD50			
CSPG6	chondroitin sulfate proteoglycan 6 (bamacan)	141542	Hs.24485	9126			SMC3			
DDX9	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 9 (RNA helicase A, nuclear DNA helicase II; leukophysin)	142190	Hs.74578	1660						
SMC1L1	SMC1 (structural maintenance of chromosomes 1, yeast)-like 1	142301	Hs.211602	8243			SMC1			



Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
TCOF1	Treacher Collins-Franceschetti syndrome 1	144313	Hs.301266	6949	Disease mutation; Polymorphism.	1	MUC1			
NNP69	Novel Nucleolar Protein	145807	Hs.10848	9790		2	YPL217c			
MYBBP1A	MYB binding protein (P160) 1a	150094	Hs.22824	10514		1				
BLM	Bloom syndrome	159001	Hs.36820	641	Hydrolase; Helicase; ATP-binding; DNA-binding; Nuclear protein;Disease mutation.	1,2	SGS1			
NNP70	Novel Nucleolar Protein	170571	Hs.57730	9816	Transmembrane.					
TOP2A	topoisomerase (DNA) II alpha (170kD)	175103	Hs.156346	7153	Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein;Polymorphism.	1	TOP2			
TOP2B	topoisomerase (DNA) II beta (180kD)	182693	Hs.75248	7155		1	TOP2			
U5-200K	KIAA0788 protein	194480	Hs.246112	23020			BRR2			
RPA190	DKFZP586M0122 protein (pol I largest subunit)	196189	Hs.59475	25885		1	RPA190			
DSP	desmoplakin (DPI, DPII)	201360	Hs.74316	1832	Repeat; Coiled coil; Phosphorylation; Cytoskeleton;Structural protein; Alternative splicing.					
RRP5	KIAA0185 protein	211472	Hs.239499	22984	Nuclear protein; rRNA processing.	2	RRP5			
CHD4	chromodomain helicase DNA binding protein 4	217991	Hs.74441	1108	DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat;Transcription regulation; Activator; Antigen.		CHD1			
FER1L3	fer-1 (C.elegans)-like 3 (myoferlin)	229587	Hs.234680	26509						
NUMA1	nuclear mitotic apparatus protein 1	238275	Hs.301512	4926						
NNP71	Novel Nucleolar Protein	244070	12544811				YJL109c			
NNP72	Novel Nucleolar Protein	254602	Hs.18759	9875						
GCN1L1	GCN1 (general control of amino-acid synthesis 1, yeast)-like 1	266340	Hs.75354	10985			GCN1			
PRP8	U5 snRNP-specific protein (220 kD), ortholog of S. cerevisiae Prp8p	273601	Hs.181368	10594			PRP8			
ATR	ataxia telangiectasia and Rad3 related	301368	Hs.77613	545	Transferase; Kinase.		MEC1			
NNP73	mRNA down-regulated in metastasis	318421	Hs.178614	27340			YBL004W			

Continued

Table S1. Proteins Identified from Nucleoli Isolated from Cultured Human Cells (continued)

Gene Name	Description	calc. Mr	Unigene	Locus	Remarks	Code	yeast GN	Dm	Ce	Sc
MKI67	antigen identified by monoclonal antibody Ki-67	319496	Hs.80976	4288		1				
PRKDC	protein kinase, DNA-activated, catalytic polypeptide	470427	Hs.155637	5591			TOR1			
NNP74	Novel Nucleolar Protein		9690317							
NNP75	Novel Nucleolar Protein		Hs.250638				DBP2			
NNP76	Novel Nucleolar Protein		Hs.70582	79039						
NNP77	Novel Nucleolar Protein	n	Hs.91579	92856		2	IMP4			
NNP78	Novel Nucleolar Protein	n	Hs.77135	84549			MAK16			
NNP79	Novel Nucleolar Protein	n	ctg13052_9							
NNP80	Novel Nucleolar Protein	n	Hs.56043	51018						
NNP81	Surfeit 6	n	Hs.274430	6838		2				
NNP82	Novel Nucleolar Protein	n	Hs.224137	51490						

Proteins are listed in ascending order according to calculated molecular weight. Where Unigene identifiers were not available, the GenBank accession number or IPI database for genomic fragment entry was used. "Locus" indicates the gene locus identifier for NCBI genomic annotation. "Remarks" were extracted from the Swiss-prot database entry, where available, and were annotated by the authors. "Code" indicates the cellular localization (if known), where r = ribosomal, 1 = published nucleolar localization for the human protein, 2 = published nucleolar localization for protein homolog/s from other species, * = tagged in this study and shown to be nucleolar, and + = tagged in this study and not observed to be nucleolar. The colors indicated at the left of each entry relate to the classification of each protein for the construction of the pie chart in Figure 5. For those proteins that have clear yeast homologs, the yeast gene names are listed (these were extracted from the Yeast Proteome database). The shaded panels represent the results of BLASTp searches for each protein against three NCBI databases: *D. melanogaster* (Dm), *C. elegans* (Ce), and *S. cerevisiae* (Sc). The results are shaded such that a black panel represents a blastp score of greater than 1000 and thus indicates a significant match; 75% grey indicates a score of between 750 and 1000; 50% grey indicates a score of between 500 and 750; 25% grey indicates a score of between 250 and 500; and a white panel indicates a score below 250. Further information on the nucleolar proteome is available at http://www.dundee.ac.uk/lifesciences/lamonddatabase/. The IPI.1 identifier numbers ctg17711_56, ctg18647_6, ctg14521_16, ctg16922_16, and ctg595_12 correspond to the IPI.2 identifier numbers IPI00019320, IPI00013458, IPI00007953, IPI00027831, and IPI00005832, respectively. ctg13052_9 is homologous to the mouse sequence AK002985.

Current Biology, Vol. 12, 13–25, January 8, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0960-9822(01)00632-7

Paraspeckles: A Novel Nuclear Domain

Archa H. Fox,[1] Yun Wah Lam,[1] Anthony K.L. Leung,[1] Carol E. Lyon,[1] Jens Andersen,[2] Matthias Mann,[2,3] and Angus I. Lamond[1,4]

[1]Wellcome Trust Biocentre
MSI/WTB Complex
University of Dundee
Dundee DD1 4HN
United Kingdom

[2]Department of Biochemistry and Molecular Biology
University of Southern Denmark
Campusvej 55
DK-5230 Odense M

[3]MDS Proteomics
Staermosegaadsvej 6
5230 Odense M
Denmark

Summary

Background: The cell nucleus contains distinct classes of subnuclear bodies, including nucleoli, splicing speckles, Cajal bodies, gems, and PML bodies. Many nuclear proteins are known to interact dynamically with one or other of these bodies, and disruption of the specific organization of nuclear proteins can result in defects in cell functions and may cause molecular disease.

Results: A proteomic study of purified human nucleoli has identified novel proteins, including Paraspeckle Protein 1 (PSP1) (see accompanying article, this issue of *Current Biology*). Here we show that PSP1 accumulates in a new nucleoplasmic compartment, termed paraspeckles, that also contains at least two other protein components: PSP2 and p54/nrb. A similar pattern of typically 10 to 20 paraspeckles was detected in all human cell types analyzed, including primary and transformed cells. Paraspeckles correspond to discrete bodies in the interchromatin nucleoplasmic space that are often located adjacent to splicing speckles. A stable cell line expressing YFP-PSP1 has been established and used to demonstrate that PSP1 interacts dynamically with nucleoli and paraspeckles in living cells. The three paraspeckle proteins relocalize quantitatively to unique cap structures at the nucleolar periphery when transcription is inhibited.

Conclusions: We have identified a novel nuclear compartment, termed paraspeckles, found in both primary and transformed human cells. Paraspeckles contain at least three RNA binding proteins that all interact dynamically with the nucleolus in a transcription-dependent fashion.

[4]Correspondence: a.i.lamond@dundee.ac.uk

Introduction

The nucleus is a complex organelle with an internal structure and component organization that is not fully characterized [1, 2]. Within the nucleus, protein and RNP complexes can move to and from chromatin through the interchromatin nucleoplasmic space, primarily by a diffusion mechanism [3, 4]. In addition, many nuclear factors are able to form or interact with different classes of subnuclear bodies, which are either attached to chromatin or located in the interchromatin space. These nuclear bodies, or "compartments," include nucleoli, splicing speckles, Cajal bodies, gems, and PML bodies [5]. They are not enclosed by membranes, yet are distinct from the surrounding nucleoplasm and contain proteins involved in different nuclear activities, such as ribosome biogenesis, transcription, and RNA splicing. Further, many observations indicate that the accurate maintenance of subnuclear organization has important consequences for cellular function [6–14].

There is increasing evidence that the internal organization of the nucleus and many of the protein and RNP factors within it are highly dynamic. The recent use of photobleaching techniques to analyze GFP-tagged proteins expressed in live cells has shown that nuclear proteins can diffuse rapidly and traffic at high rates between separate compartments [15–18]. The nuclear localization pattern of many proteins and, as a consequence, the overall structure of the nucleus has been shown to change during the cell cycle and in response to changes in cellular metabolic activity. For example, inhibiting transcription with drugs such as Actinomycin D, 5,6-dichloro-1-β-D-ribofuranosylbenzimadizole (DRB), or α amanitin all cause a significant redistribution of many nuclear factors and alter the morphologies of nucleoli and splicing speckles [19, 20].

Recent studies have started to provide clues concerning the specific roles of individual forms of nuclear bodies. The interchromatin granule clusters (speckles) that contain snRNPs and other spliceosome components may be sites for storage or recycling of splicing factors, although in some cases they may also be more directly involved in the splicing mechanism [21–23]. Cajal bodies specifically accumulate snRNPs and nucleolar snoRNPs that have recently been imported into the nucleus and are likely undergoing maturation [24–27]. Cajal bodies may have a role in RNA modification and transport and may also be involved in the feedback regulation of gene expression from snRNA gene loci [28]. In the case of nucleoli, it is known that they are sites for the transcription of rRNA genes and the subsequent processing of rRNA precursors and assembly of ribosomal subunits. Interestingly, there is also evidence that nucleoli may perform additional nuclear functions, such as controlling the activity of regulatory molecules by a sequestration mechanism and participating in the biogenesis of other RNP complexes [29–31].

The ability to purify large quantities of relatively pure nucleoli makes it an ideal candidate for biochemical

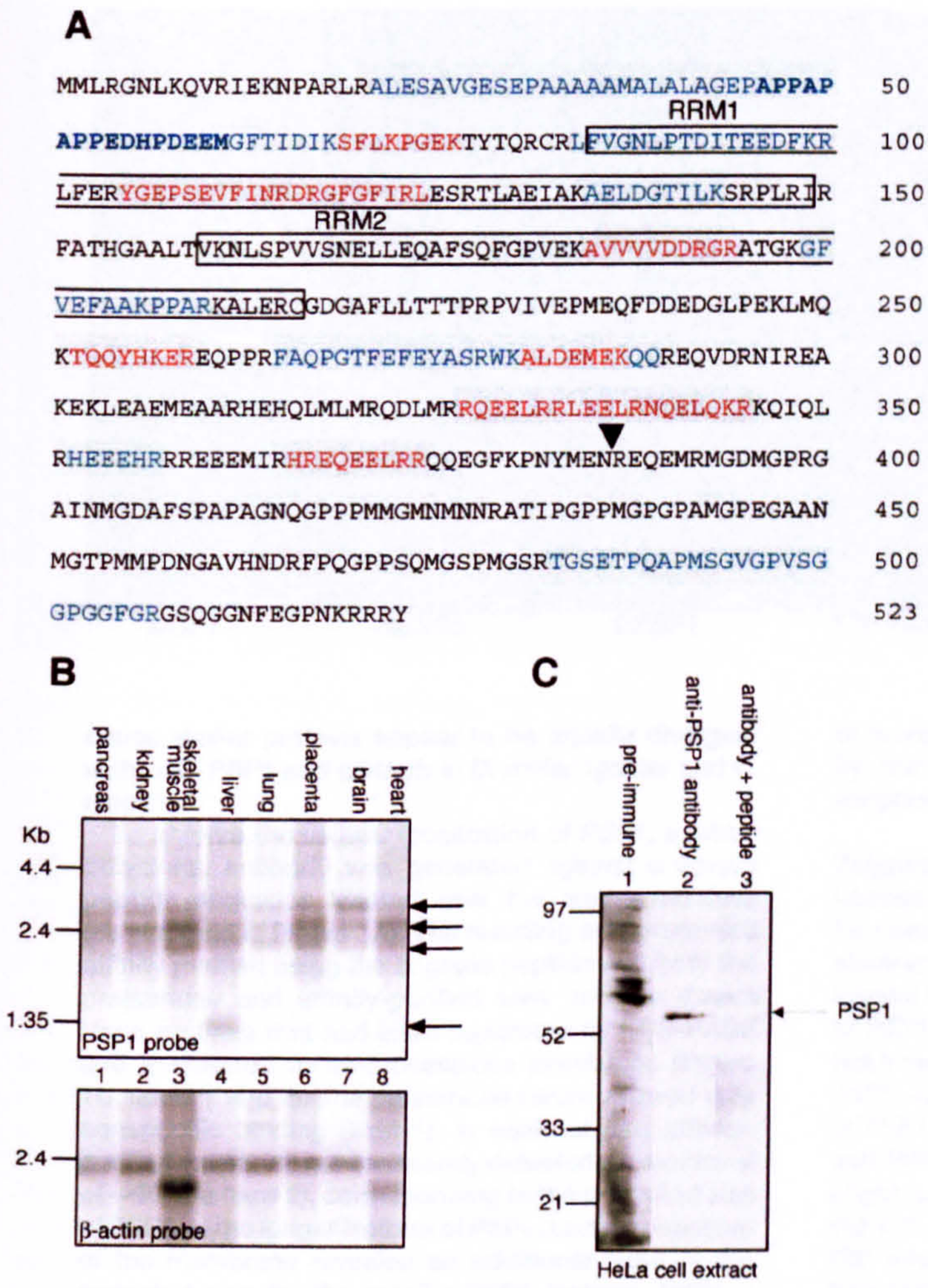


Figure 1. The Sequence and Expression of Paraspeckle Protein 1

(A) Amino acid sequence of PSP1. Prepublication accession numbers for PSP1: Genbank FLJ10955 and ENSEMBL gene ID ENSG00000121390. The arrowhead after residue 386 indicates the final common residue for both splice forms of PSP1. The shorter form (PSP1-β) has an additional 7 residues (GDKRKCG, a translation of exon 7), making it a total of 393 residues long. The remaining residues C-terminal to the arrowhead are the extent of the longer splice variant PSP1-α (523 residues long), composed of exons 8 and 9 as defined by the genomic annotation. Red and blue text indicates tryptic peptides identified using MS on PSP1 copurified with HeLa nucleoli. For further details of PSP1 splice forms, see Supplementary Material. The peptide at the N-terminus of PSP1 that the PSP1 antibody was generated against is shown in bold; this sequence is in a region of the protein with low similarity to p54/nrb and is thus unique to PSP1. The RNA-binding domains are boxed.

(B) Human Northern blot probed for PSP1. A human multitissue Northern blot (Clontech) was hybridized with a ^{32}P-radiolabeled cDNA probe for PSP1 (see Experimental Procedures) (upper panel), then stripped and reprobed with a β-actin probe to show approximately equal loading of mRNA in each lane (lower panel), as per the manufacturer's instructions. The hybridized radioactivity was visualized using a Phosphoimager. Tissue sources for the mRNA in each lane are indicated on the figure. We note that the signal for the PSP1 probe was considerably weaker than for the β-actin probe, possibly indicating that this is a low-abundance transcript.

(C) PSP1 protein in HeLa extract. HeLa cells were lysed and the resultant extract subjected to SDS/PAGE and transfer to nitrocellulose. The blot was divided into three and probed as follows: rabbit preimmune serum (PI) (lane 1), anti-PSP1 antiserum (lane 2), or anti-PSP1 antiserum preincubated with a 10 molar excess of cognate peptide (lane 3). These treatments were followed by incubation with HRP-conjugated anti-rabbit antibody and detection by chemiluminescence using ECL plus substrate (Amersham Pharmacia Biotech). The arrow indicates the major band of ~60 kDa in size that is detected specifically by the anti-PSP1 antiserum.

analysis in vitro. A major proteomic analysis of nucleoli isolated from cultured human cells has therefore been undertaken to characterize the protein composition of this important subnuclear body [32]. This study identified 271 human nucleolar proteins using MS, including 80 proteins encoded by novel or uncharacterized human genes. Here we report the characterization of one of the novel human nucleolar factors—Paraspeckle Protein 1 (PSP1). We show that PSP1 accumulates within a new form of nuclear compartment in the interchromatin space, termed paraspeckles. This analysis reveals a new level of structural complexity within the nucleus and highlights the dynamic nature of nuclear body proteins.

Results

PSP1: A Novel Protein from the Human Nucleolar Proteome that Localizes to a Punctate Pattern in the Nucleus

PSP1 was initially chosen for detailed analysis because we observed that it belonged to a subset of human nucleolar proteins that were enriched in nucleoli isolated from cells treated with the transcription inhibitor Actinomycin D [32]. Based upon a combination of cDNA sequence analysis and direct mass spectrometric identification of tryptic peptides, we deduce that the major form of PSP1 is a basic protein of 523 amino acids with a predicted Mr of 58.7 kDa (Figure 1A). As judged by probing a human multitissue Northern blot, PSP1 is encoded by several ubiquitously expressed mRNAs of ~2.5 kb, including alternatively spliced forms (Figure 1B). We have named the two major expressed isoforms PSP1-α and PSP1-β (see Figure 1 and Supplementary Material available with this article online). The primary amino acid sequence of PSP1 contains two conserved RNA binding domains (RRMs) located at the amino terminus of both isoforms and is most closely related to the nuclear protein p54/nrb [33, 34]. PSP1 is highly conserved in mammals, with a mouse ortholog showing greater than 95% identity at the amino acid level (Genpept accession number BAB27509). However, in lower organisms it is difficult to define clear orthologs, be-

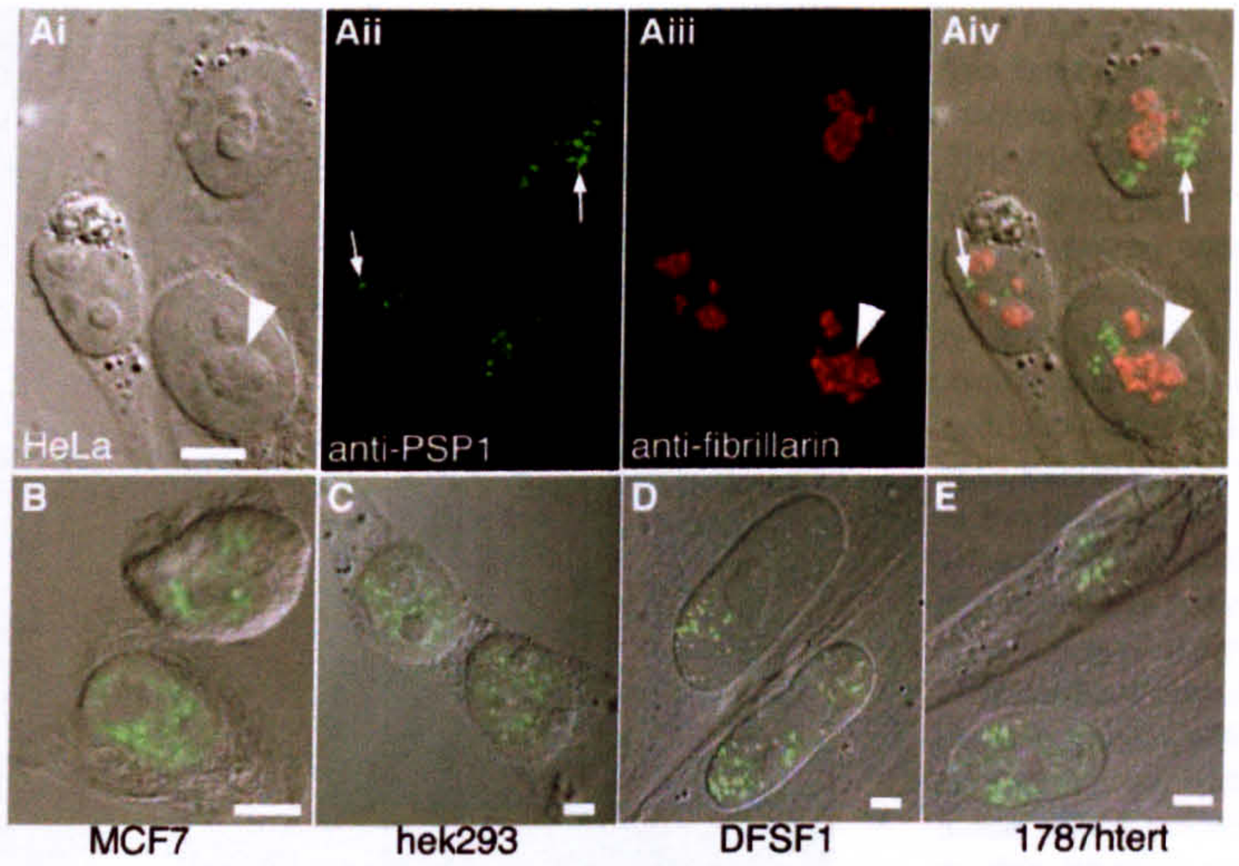


Figure 2. PSP1 Has a Punctate Nuclear Localization in Human Cells

The confocal fluorescence micrographs show PSP1 antiserum staining in the transformed cell lines (A) HeLa, (B) MCF7, (C) Hek293, and in the "primary" (D) DFSF1 and (E) 1787htert cells. An overlay of the Nomarski image for each cell type and the PSP1 immunostaining is shown in (B)–(E). The HeLa cells in (A) have also been immunostained using anti-fibrillarin antiserum (shown in panel [Aiii]), and the overlay of the Nomarski image for the cells as well as the fibrillarin and PSP1 signals is shown in panel (Aiv). Each fluorescent image is a stack of confocal sections. Arrowheads indicate nucleoli as seen in Nomarski and fibrillarin staining, and arrows indicate punctate PSP1 spots. Scale bars, 5 μm.

cause similar proteins appear to be equally divergent with both PSP1 and p54/nrb in *D. melanogaster* and *C. elegans*.

To analyze the cellular localization of PSP1, a rabbit polyclonal antibody was generated against a unique peptide sequence located near the amino terminus (shown bold in Figure 1A). The resulting antiserum was affinity purified using the cognate peptide and both the preimmune and affinity-purified sera used to detect HeLa proteins that had been separated by SDS-PAGE and transferred to a nitrocellulose membrane (Figure 1C, lanes 1 and 2). The preimmune serum showed only nonspecific binding (lane 1). In contrast, the affinity-purified antiserum predominantly detected a major band of ~60 kDa (lane 2), corresponding to the predicted size of PSP1-α, the longer isoform of PSP1. Longer exposure of the membrane revealed an additional band of the expected size for the smaller PSP1 isoform, PSP1-β (data not shown). The specific PSP1 bands were no longer detected when the affinity-purified antiserum was preblocked with the PSP1 peptide (lane 3).

Having established the specificity of the anti-PSP1 serum, we used it for indirect immunofluorescence staining of HeLa cells (Figure 2A). The anti-PSP1 antiserum specifically labeled HeLa nuclei in a punctate pattern, with some diffuse nucleoplasmic labeling (arrows in Figures 2Aii and 2Aiv). Surprisingly, although PSP1 was identified through proteomic analysis of nucleoli, the anti-PSP1 antiserum did not label nucleoli significantly, as shown by double-labeling experiments using an anti-fibrillarin antibody to detect nucleoli (Figures 2Ai, 2Aiii, and 2Aiv). The anti-PSP1 antiserum typically labeled instead 10 to 20 bright punctate structures scattered throughout the nucleoplasm.

The anti-PSP1 antiserum also labeled a similar pattern of punctate, nucleoplasmic structures, with varying levels of diffuse nucleoplasmic labeling but no significant nucleolar staining, in both primary and transformed human cell lines (Figures 2B–2E). These included transformed lines derived from either breast (MCF7, Figure 2B) or kidney cells (hek293, Figure 2C) and primary cells corresponding to foreskin fibroblasts (DFSF1, Figure 2D) or human fibroblasts extended in proliferative life span by retroviral expression of telomerase reverse transcriptase (1787htert, Figure 2E).

Tagged PSP1 Shows the Same Punctate Localization as Endogenous PSP1

To confirm that the punctate nuclear staining pattern observed with anti-PSP1 antiserum in primary and transformed human cells genuinely reflected the localization of PSP1, we transiently expressed exogenous PSP1 in HeLa cells fused with either a yellow fluorescent protein (YFP), or myc peptide tag (Figure 3). Both tagged forms of PSP1 showed a similar pattern to that seen with the anti-PSP1 antiserum in untransfected cells, including bright punctate nuclear structures. The tagged proteins did not accumulate in nucleoli, as detected by either DIC microscopy or immunolabeling with an anti-fibrillarin antibody (Figures 3A and 3B; arrowheads denote nucleoli; arrows denote tagged PSP1).

We next established stable HeLa cell lines expressing YFP-PSP1 (Figure 4). Separate cell lines were isolated for the two predominant splice forms of PSP1. As both isoforms gave identical results, subsequent data are presented only for the PSP1-β isoform. Immunolabeling both the parental and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell lines with anti-PSP1 antiserum detected a similar pattern of punctate, nucleoplasmic structures and diffuse nucleoplasmic staining (Figure 4, cf. panels 4Aiii and 4Biii). Analysis of 50 labeled cells from both the parental and stable cell lines showed that there was no statistical difference in the average number of punctate PSP1 structures labeled with PSP1 antiserum or YFP fluorescence (~15 per nucleus, data not shown; for example, Figure 4, panels 4Bii and 4Biii).

Proteins from both the parental and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells were separated by SDS-PAGE, transferred to nitrocellulose membrane, and probed with either anti-GFP (Figure 4C, lanes 1 and 2) or anti-PSP1 antiserum (Figure 4C, lanes 3–6). A single band was detected with the anti-GFP antibody specifically in the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells (lane 2). In contrast, the major endogenous PSP1 band was detected in both cell lines using the anti-PSP1 antiserum

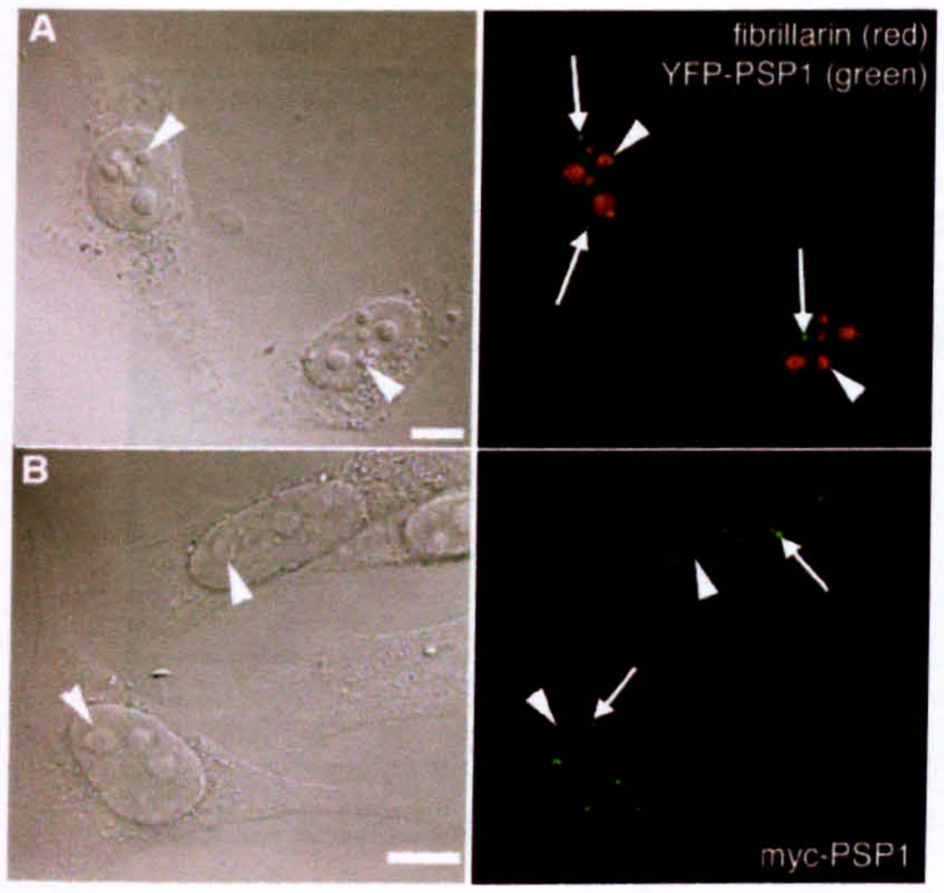


Figure 3. Tagged PSP1 Localizes to Punctate Foci in HeLa Nuclei

Confocal fluorescence micrographs are shown on the right, with the corresponding Nomarski images to the left. HeLa cells were transiently transfected with YFP-PSP1 (A) or myc-PSP1 (B) and fixed 16 hr after transfection. Panel (A) shows direct fluorescence of the YFP-PSP1 protein (green) and immunofluorescence staining for fibrillarin (red). Panel (B) shows HeLa nuclei immunostained with anti-myc antiserum (green). Large arrowheads indicate nucleoli, and arrows indicate the punctate YFP-PSP1 and myc-PSP1. Scale bars, 5 μm.

(lanes 3 and 4) but was absent when the antiserum was preblocked with the cognate peptide (lanes 5 and 6). We observe that the YFP-PSP1 fusion protein is expressed at ~4-fold higher levels than the endogenous PSP1 protein in HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells (Figure 4C, lane 4 and other data not shown). However, the relative level of endogenous PSP1 appears higher in the parental cell line, suggesting that some degree of downregulation of endogenous PSP1 may have occurred to compensate for expression of the fusion protein (Figure 4C, cf. lanes 3 and 4).

We compared the physical properties of the endogenous and YFP-tagged PSP1 proteins in the parental and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell lines (Figure 4D). Cultured cells were extracted sequentially with buffers containing varying concentrations of salt and detergent, and the resulting protein samples were separated by SDS-PAGE, transferred to nitrocellulose, and probed with anti-PSP1 antiserum. In both the parental and stable cell lines, PSP1 was quantitatively recovered in the second fraction (Figure 4D, lanes 2 and 6). Furthermore, both the YFP-tagged and endogenous PSP1 were recovered in the same fraction from the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells (Figure 4D, lane 6). Thus, the tagged and endogenous PSP1 proteins have similar properties. Further comparison of the parental and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell lines showed no apparent change in their respective viability or rate of cell division (data not shown). We therefore conclude that the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell line is viable and expresses a single tagged YFP-PSP1 fusion protein that has a similar localization pattern and physical properties to the endogenous PSP1.

PSP1 Localizes in a Novel Nuclear Compartment

We next sought to identify which of the known subnuclear compartments PSP1 was accumulated in. Therefore, double-labeling experiments were carried out using the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell line to detect PSP1 and immunolabeling with characterized antisera to detect marker proteins for separate nuclear bodies (Figures 5A–5D). Antibodies specific for snRNPs (anti-2,2,7-trimethylguanosine-cap structure) and for the non-snRNP splicing factor SC35 were both used to detect splicing speckles (Figures 5A and 5B). We observed no overlap in the localization of PSP1 and splicing speckles using either of these antibodies (Figures 5Aiii and 5Biii; in each panel, arrows indicate YFP-PSP1, and arrowheads indicate known nuclear markers, e.g., speckles). The brighter foci detected with the snRNP marker also did not colocalize with YFP-PSP1 (Figures 5Aii and 5Aiii, broken arrows indicate bright foci). It is most likely that these bright snRNP foci represent Cajal bodies [20]. A double-labeling experiment with anti-coilin antiserum (a marker for Cajal bodies) confirmed that there is no overlap or colocalization with YFP-PSP1 (Figure 5Ciii). We also detected no colocalization of YFP-PSP1 either with the bodies labeled using an anti-SMN antiserum, which detects both Cajal bodies and gems [9] (data not shown), or with an anti-U1A antibody, which should label interchromatin granule associated zones [35] as well as speckles (data not shown). The other major class of subnuclear bodies that have been studied contain the cellular RING finger protein PML [6–8, 36]. Double labeling using an anti-PML antiserum showed that the YFP-PSP1 signal also does not localize or associate with PML bodies (Figure 5Diii). We next compared the localization of YFP-PSP1 with the major sites of RNA synthesis, as detected by pulsed incorporation of Br-UTP (Figure 5E). Despite the extensive nucleoplasmic labeling detected with the anti-BrdUTP antiserum, we again observed no obvious colocalization with the punctate YFP-PSP1 signal (Figure 5Eiii).

Finally, we used a triple-labeling protocol to compare the localization of YFP-PSP1 with both bulk chromatin, as detected by DAPI staining, and splicing speckles, as detected using anti-SC35 antiserum (Figure 6). This showed that the YFP-PSP1 localized predominantly to interchromatin regions that have little or no DAPI signal (Figure 6, DAPI-poor regions indicated by arrowheads). A similar result was obtained using a fluorescently tagged histone H2B and the anti-PSP1 antiserum (data not shown). Interestingly, although we still detected no colocalization of YFP-PSP1 with the SC35 domains, which also localize to the interchromatin regions, analysis of optical z sections of the triple-labeled nuclei showed that the YFP-PSP1 structures were frequently juxtaposed beside SC35 speckles (Figure 6, lower panels; arrows indicate YFP-PSP1, and arrowheads indicate SC35 speckles in DAPI-poor regions).

In conclusion, PSP1 concentrates in a novel form of punctate subnuclear structure that localizes to the interchromatin space. Considering both the speckled appearance of the PSP1-containing structures and their frequent localization adjacent to SC35 speckles, we have termed this novel compartment "paraspeckles."

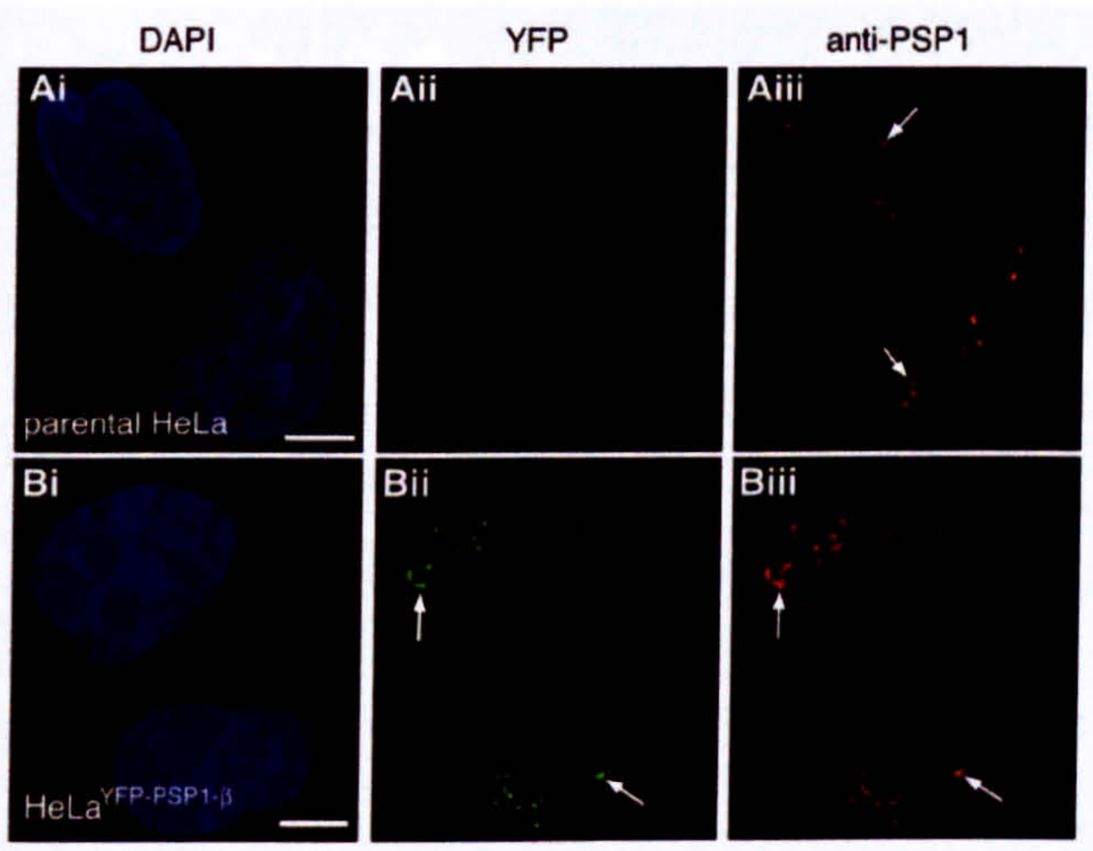


Figure 4. Stable Expression of YFP-PSP1 in HeLa Cells Results in the Same Punctate Pattern of PSP1 Localization in the Nucleus

Fluorescent micrographs of HeLa (A) and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ (B). The extent of the nucleus is shown with DAPI staining in panels (Ai)–(Bi), the fluorescence from the exogenously expressed YFP-PSP1 in panels (Aii)–(Bii), and the staining pattern for endogenous PSP1 as determined by staining with anti-PSP1 antiserum in panels (Aiii)–(Biii). All images are maximum intensity projections of stacks deconvolved using the deltavision Softworx software. Arrows point to examples of the same foci detected with the PSP1 antibody as appear in the YFP fluorescence channel. Scale bars, 5 μm. (C) Anti-PSP1 antiserum recognizes both the endogenous and fusion PSP1 proteins, while anti-GFP antiserum only recognizes YFP-PSP1. Equal amounts of HeLa cell extract (lanes 1, 3, and 5) and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ stable cell extract (lanes 2, 4, and 6) were electrophoresed, transferred to nitrocellulose, then immunoblotted with anti-GFP antibody (lanes 1 and 2), anti-PSP1 antibody (lanes 3 and 4), or anti-PSP1 antibody preblocked by incubation with 10-fold excess of cognate peptide (lanes 5 and 6). An anti-rabbit peroxidase conjugated secondary antibody was employed to detect antibody binding, and this in turn was detected using ECL plus (Amersham Pharmacia Biotech). (D) Biochemical behavior of endogenous PSP1 and YFP-PSP1. HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells (samples in lanes 5–8) and parental HeLa cells (samples in lanes 1–4) were both extracted in several buffers containing different concentrations of salts and detergents (see Experimental Procedures). Equivalent protein samples from each sequential extract were loaded onto a gel, electrophoresed, transferred to nitrocellulose, and immunoblotted with anti-PSP1 antibody. The following extracts were loaded in each lane: supernatant A (lanes 1 and 5), supernatant B (lanes 2 and 6), supernatant C (lanes 3 and 7), and pellet C (lanes 4 and 8). Both endogenous and YFP-PSP1 are specifically released from the YFP-PSP1 cells in supernatant B (lane 6), and the endogenous PSP1 from parental HeLa is also specifically released in supernatant B (lane 2). Arrows indicate endogenous PSP1 and exogenous YFP-PSP1 proteins.

PSP1 Relocalizes to the Nucleolus When Transcription Is Blocked

We next examined how the localization of PSP1 was affected by ongoing gene expression, which is known to alter the structure of other subnuclear bodies. After treatment of the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell line with Actinomycin D, which halts elongation by RNA polymerases, we observed a dramatic reorganization of PSP1 from a pattern of paraspeckles to discrete caps formed at the nucleolar periphery (Figure 7A, cf. panels 7Ai and 7Aii). This response contrasts with the behavior of splicing speckle proteins, which concentrate in enlarged speckles in the nucleoplasm after Actinomycin D treatment, as shown here for SC35 (Figure 7Aiii and 7Aiv; arrows indicate PSP1, and arrowheads indicate SC35). A similar effect was observed upon treating cells with a separate inhibitor of RNA synthesis, DRB, indicating that it most likely results from an inhibition of transcription (data not shown). Analysis of PSP1 localization in a live HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell at multiple time points after addition of Actinomycin D to the culture medium showed a progressive loss of PSP1 signal from nucleoplasmic paraspeckles and a concomitant appearance of PSP1 signal at the nucleolus (Figure 7B). We never observed the direct movement of whole paraspeckles to the nucleolus, suggesting that the PSP1 protein can flux out of the paraspeckles and accumulate at a separate nuclear location when transcription is inhibited.

The data above are consistent with the proteomic analysis of nucleoli, which showed that a subset of 11 proteins, including PSP1, was enriched in nucleoli following treatment of cells with Actinomycin D [32]. We asked if any of these 11 proteins shared a paraspeckle localization with PSP1. However, for most of these enriched proteins, previous studies have reported that they localize to known nuclear structures [37–40]. For example, our analysis of YFP-tagged p68 and p72 showed that they did not localize to paraspeckles (data not shown) but rather displayed a diffuse punctate nucleoplasmic pattern as previously described [38]. However, transient expression of YFP-tagged p54/nrb, the most closely related PSP1 homolog, showed that it was also a paraspeckle protein as judged by double labeling with the anti-PSP1 antiserum (data not shown). Interestingly, another novel protein identified in the proteomic study

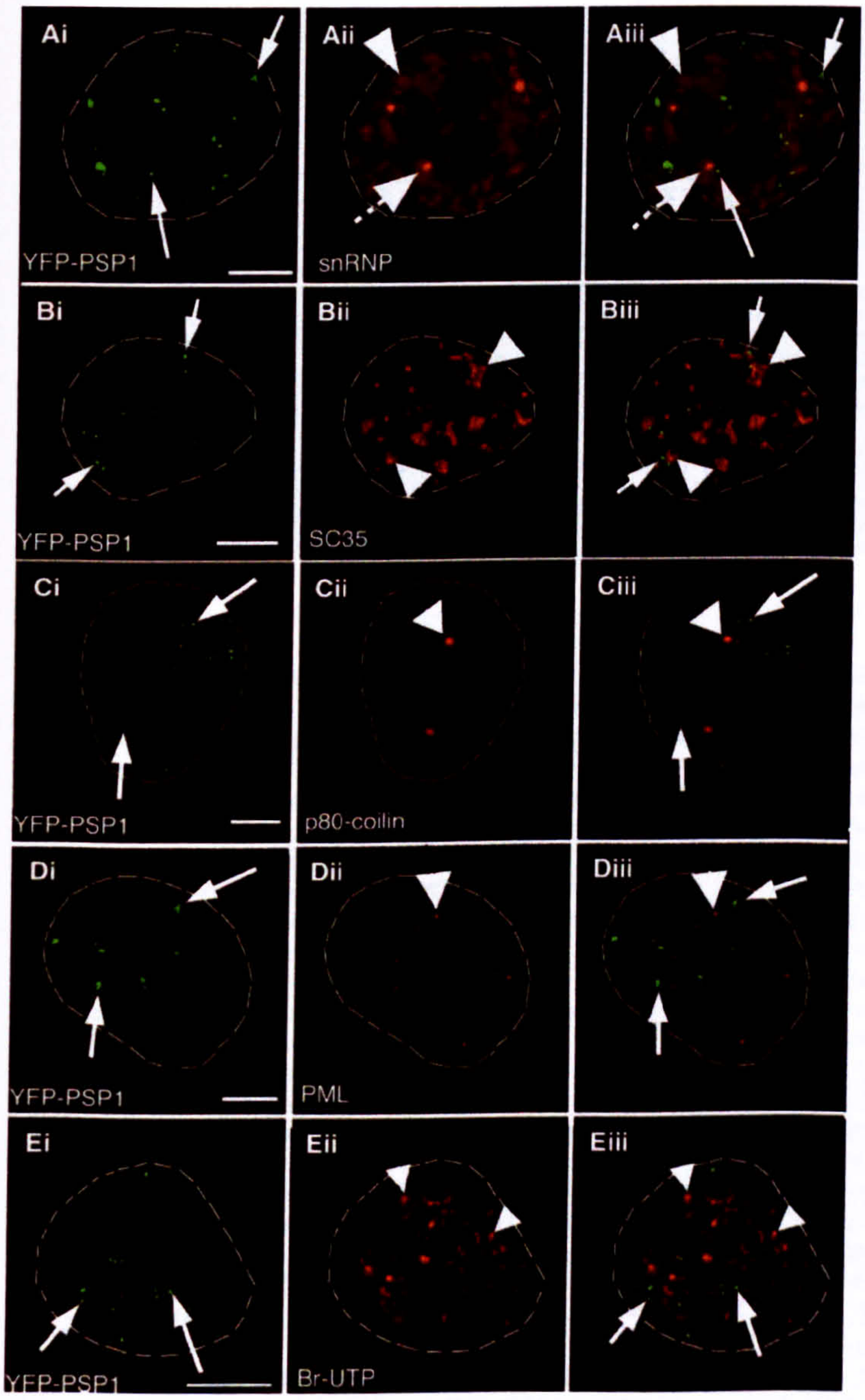


Figure 5. PSP1 Localizes to a Novel Subnuclear Compartment

HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells were fixed and immunostained with antibodies against nuclear markers (red, panels [Aii]–[Dii]), the YFP-PSP1 fluorescence is shown in green in panels (Ai)–(Ei), and the overlay in panels (Aiii)–(Eiii). The following proteins were labeled: (A) snRNP splicing speckles as defined by staining with anti-5′-trimethylguanosine antiserum, (B) SC35 splicing speckles, (C) p80-coilin labeling Cajal bodies, and (D) PML labeling PML bodies. In (E), the transcription sites in HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells were labeled using Br-UTP incorporation, followed by detection with anti-BrdUTP antiserum (panel [Eii]), and the overlay is shown in panel (Eiii). Each panel shows maximum intensity projections of deconvolved sections acquired using a deltavision system. The nuclei are outlined with hand-drawn dashed lines to indicate the nuclear periphery. Scale bars, 5 μm. Large arrowheads indicate the known nuclear structures, broken arrows indicate Cajal bodies in panels (Aii) and (Aiii), and small arrows indicate YFP-PSP1 foci.

as enriched in nucleoli from Actinomycin D-treated cells also localized to paraspeckles and therefore has been named "PSP2" (Figure 7C). Unlike p54/nrb, PSP2 shows minimal homology to PSP1, although it also contains two amino-terminal RNA binding domains. Transient expression of YFP-tagged PSP2 and double labeling with anti-PSP1 antiserum revealed that it accumulates in paraspeckles and shows diffuse nuclear staining and, like PSP1, also relocalizes to perinucleolar caps following Actinomycin D treatment (Figure 7C). Furthermore, all three paraspeckle proteins, i.e., PSP1, PSP2, and p54/nrb, colocalize in the same perinucleolar caps after Actinomycin D treatment (data not shown).

Other nuclear proteins, including fibrillarin and p80 coilin, are also known to relocalize into perinucleolar caps following treatment of cells with transcription inhibitors [20, 32, 41, 42]. We therefore performed triple-labeling experiments in Actinomycin D-treated HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells with anti-p80 coilin and anti-fibrillarin antibodies (Figure 7D). This showed that all three proteins localize to different and distinct cap structures at the nucleolar periphery (Figure 7Div).

The twin observations that PSP1 can be detected, albeit in low amounts, in purified nucleoli and that it becomes enriched in nucleoli after Actinomycin D treatment ([32] and this study, Figure 7A), suggest the possibility that it cycles continually between paraspeckles and the nucleolus. To add support to this hypothesis,

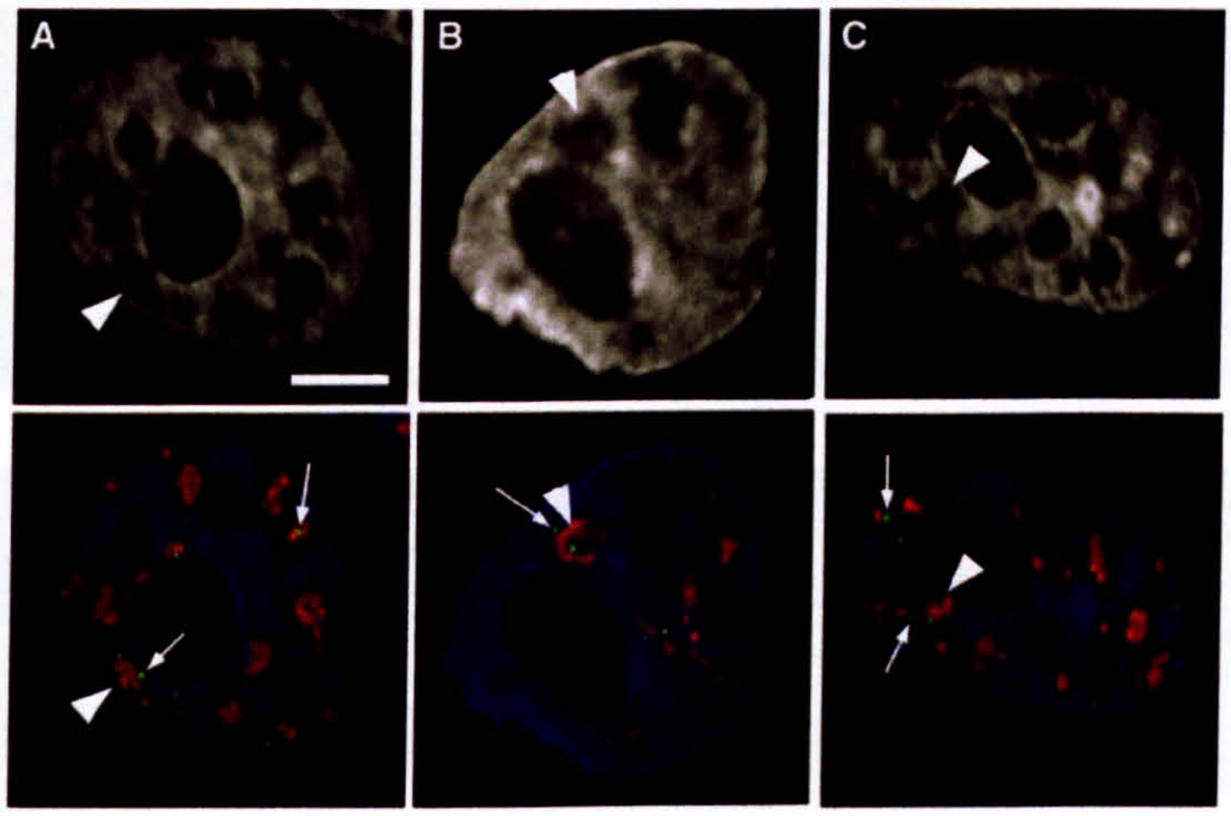


Figure 6. Paraspeckles Are Found in Interchromosomal Spaces, Often Next to SC35 Speckles

Upper panels indicate single sections through three representative HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ nuclei (A–C) stained with DAPI to highlight chromatin. The arrowheads indicate the DAPI-poor staining interchromosomal regions. The lower panels show DAPI staining (blue), YFP-PSP1 fluorescence (green), and SC35 immunostaining (red) from the same sections from the same nuclei as shown in the upper panels. The sections were obtained using the Deltavision system and the images were deconvolved using Softworx. Small arrows indicate paraspeckles, and arrowheads indicate SC35 clusters that reside within the dark, low-stained DAPI areas. Scale bar, 5 μm.

we performed a fluorescence loss in photobleaching (FLIP) experiment, using the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell line, to see if photobleaching of the nucleolus affects the fluorescence signal of YFP-PSP1 in paraspeckles (Figure 8). Several cells were imaged over the course of the experiment, while pulse photobleaching a region of interest inside the nucleolus of one cell only. The nucleoplasmic fluorescence of the cell in which the nucleolus was bleached decreased over the time course of the experiment, as compared with the unbleached cell (Figure 8A). This indicates that nucleoplasmic PSP1 was moving into the nucleolus where it became subject to bleaching. We measured the fluorescence intensity of the individual paraspeckles within the bleached cell and compared this with paraspeckles from an adjacent, unbleached control cell (Figure 8B). This showed a sequential loss of fluorescence in the paraspeckles of the cell in which the nucleolus was bleached. We conclude that the most likely explanation of this result is that PSP1 cycles between the paraspeckles and nucleolus.

Discussion

We have identified in this study a novel nuclear compartment (paraspeckles) and shown that at least three nuclear proteins, i.e., PSP1, PSP2, and p54/nrb, accumulate in these structures. The paraspeckles were detected in both primary and transformed human cell lines and are different from other well-characterized types of subnuclear bodies. They are dynamic structures containing proteins that can cycle to and from nucleoli. The identification of paraspeckles underlines the high degree of organization of nuclear factors and the potential to segregate proteins within the nucleoplasm into biochemically and spatially distinct structures.

Considering that PSP1 was first identified using MS as a nucleolar factor [32], it was initially surprising that localization studies did not detect it accumulated in nucleoli. However, as reported above, it is likely that PSP1 cycles between paraspeckles and nucleoli, thus explaining its detection in nucleolar preparations by MS. The fact that proteins that may interact dynamically or transiently with nucleoli were nonetheless detected by MS analysis illustrates the utility of large-scale proteomic analyses of purified cellular structures. Furthermore, the MS data showing that PSP1, PSP2, and p54/nrb are all members of a group of proteins enriched in nucleoli isolated from Actinomycin D-treated cells also supports the fluorescence microscopy data in this study, showing that these proteins accumulate at the nucleolar periphery when cells are exposed to Actinomycin D. While we have identified here three separate paraspeckle proteins, it may in the future be possible to use a proteomic approach to characterize the full range of proteins localized in paraspeckles if a suitable isolation procedure can be developed. In this regard, the recent report describing the identification of proteins associated with a purified interchromatin granule cluster (speckle) fraction from mouse liver cells [43] is encouraging for future efforts in purifying subnuclear bodies.

The present data show that the distribution patterns of the three paraspeckle proteins likely reflect steady state accumulations rather than static localizations, consistent with the dynamic behavior reported also for other nuclear factors (reviewed in [3, 4]). Using the HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell line, we observed that photobleaching of nucleoli led to a progressive loss of the YFP-PSP1 signal from paraspeckles. Several recent studies have also reported the converse phenomenon, i.e., bleaching of the nucleoplasm leading to a loss of fluorescent signal for a variety of nucleolar proteins [15–17]. These data collectively reveal that, for both factors accumulated in nucleoli and factors in nucleoplasmic structures, a high level of exchange can occur between compartments. This emphasizes the importance of dynamic studies and time-lapse analyses in order to understand the significance of localization patterns revealed by fluorescence microscopy.

Although the nucleolus is known to be the site of rRNA synthesis and ribosomal subunit assembly, it has been suggested that nucleoli can also perform other functions [30, 31, 44]. In this regard it is interesting that we have detected three proteins with no known connection to

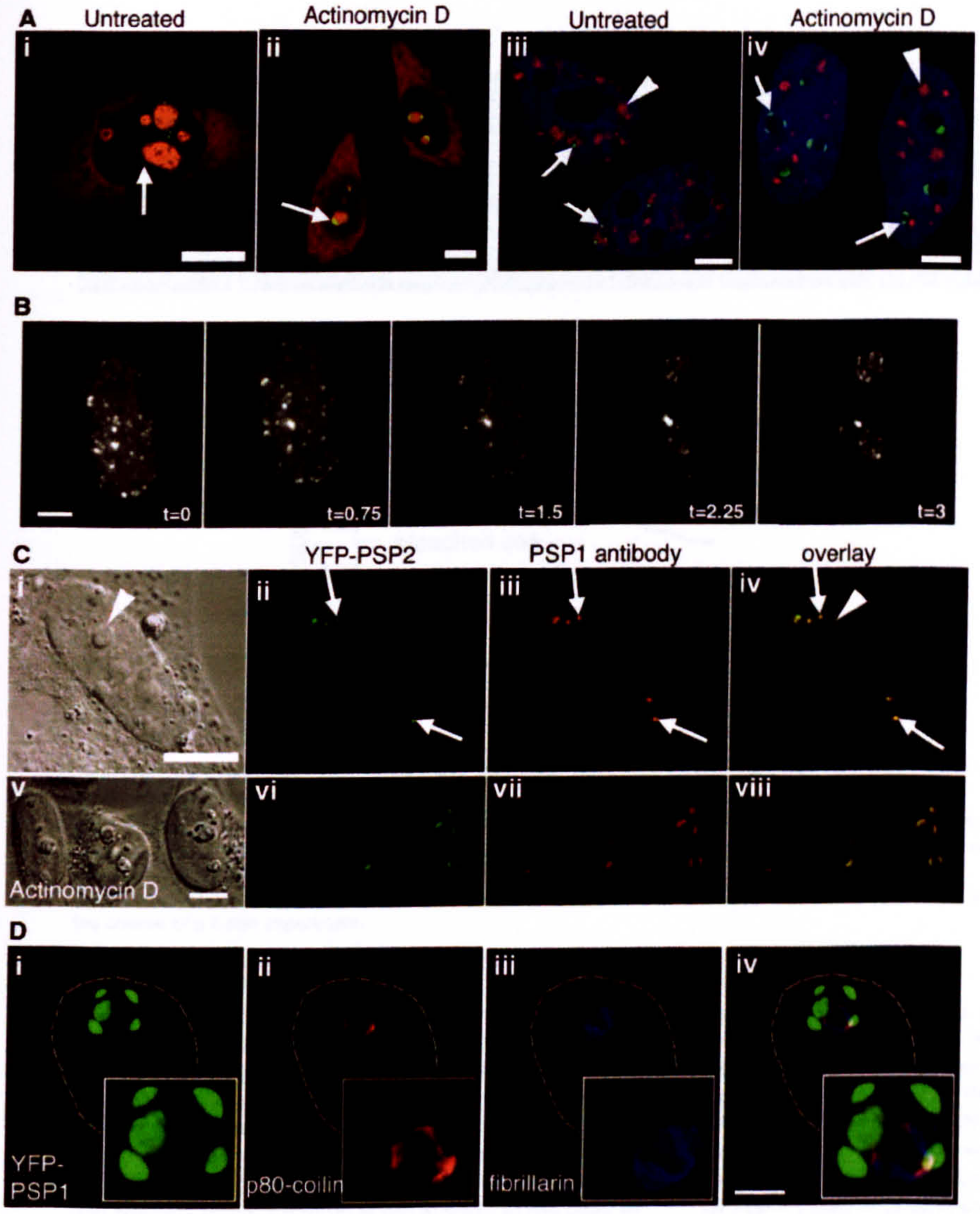


Figure 7. Inhibition of RNA Polymerase II Transcription Results in the Relocalization of PSP1 and Other Paraspeckle Proteins to the Nucleolus

(A) Fluorescence micrograph of sections through HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells untreated (Ai and Aiii) and following incubation with media containing Actinomycin D (1 μg/ml) for 4 hr (Aii and Aiv). YFP-PSP1 fluorescence is shown in green in all panels. For other colors, panels (Ai) and (Aii) show cellular RNA (including nucleoli) by staining with Pyronin Y (red), and panels (Aiii) and (Aiv) show DNA as indicated by DAPI staining (blue) and SC35 as indicated by immunostaining with anti-SC35 antiserum (red). Small arrows indicate YFP-PSP1 paraspeckles and relocalized YFP-PSP1 at the nucleolar periphery. Arrowheads indicate SC35 clusters in untreated cells and relocalized SC35 enlarged speckles after treatment with Actinomycin D.

(B) PSP1 relocalizes in response to transcription inhibition in live cells. The panels show a series of time points demonstrating the gradual relocalization of PSP1 from paraspeckles to nucleoli from a movie of a live HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell imaged following the addition of media containing 1 μg/ml Actinomycin D. The cell was imaged as described in Experimental Procedures, every 10 min over a 3 hr period.

(C) PSP2 is another paraspeckle protein. HeLa cells were transiently transfected with an expression vector for YFP-PSP2 and were fixed and immunostained using anti-PSP1 antiserum 16 hr later. Nomarski images of cells are shown in panels (Ci) and (Cv). The confocal fluorescence micrographs show YFP-PSP2 fluorescence in panels (Cii) and (Cvi), PSP1 immunostained signal in panels (Ciii) and (Cvii), and the overlays of the two channels are shown in panels (Civ) and (Cviii). Arrows indicate examples of the structures/paraspeckles that contain both PSP1 and PSP2. Panels (Ci)–(Civ) depict an untreated HeLa cell, while panels (Cv)–(Cviii) show HeLa cells that were incubated with Actinomycin D prior to fixation and immunostaining for PSP1. Note that one of the cells in panels (Cv)–(Cviii) does not have an YFP-PSP2 signal, presumably as it did not take up the expression plasmid in the transient transfection.

(D) Fibrillarin, coilin, and PSP1 are in distinct domains around the nucleolus when transcription is inhibited. Panels show a 3D reconstruction of a HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell that has been fixed and immunostained for coilin (Dii) and fibrillarin (Diii) following incubation with Actinomycin D (1 μg/ml for 4 hr). YFP-PSP1 fluorescence is shown in (Di), and the overlay of the three signals is shown in (Div), note that, due to displaying a 3D structure in 2D, some artificial overlap of the colors is apparent. The insets in each panel are shown to more clearly demonstrate the distribution of the proteins around the nucleolar space. The images are 3D volume views of a stack of 60 × 0.2 μm sections deconvolved using the Softworx package. The nucleus is outlined with a hand-drawn dashed line. Scale bars, 5 μm.

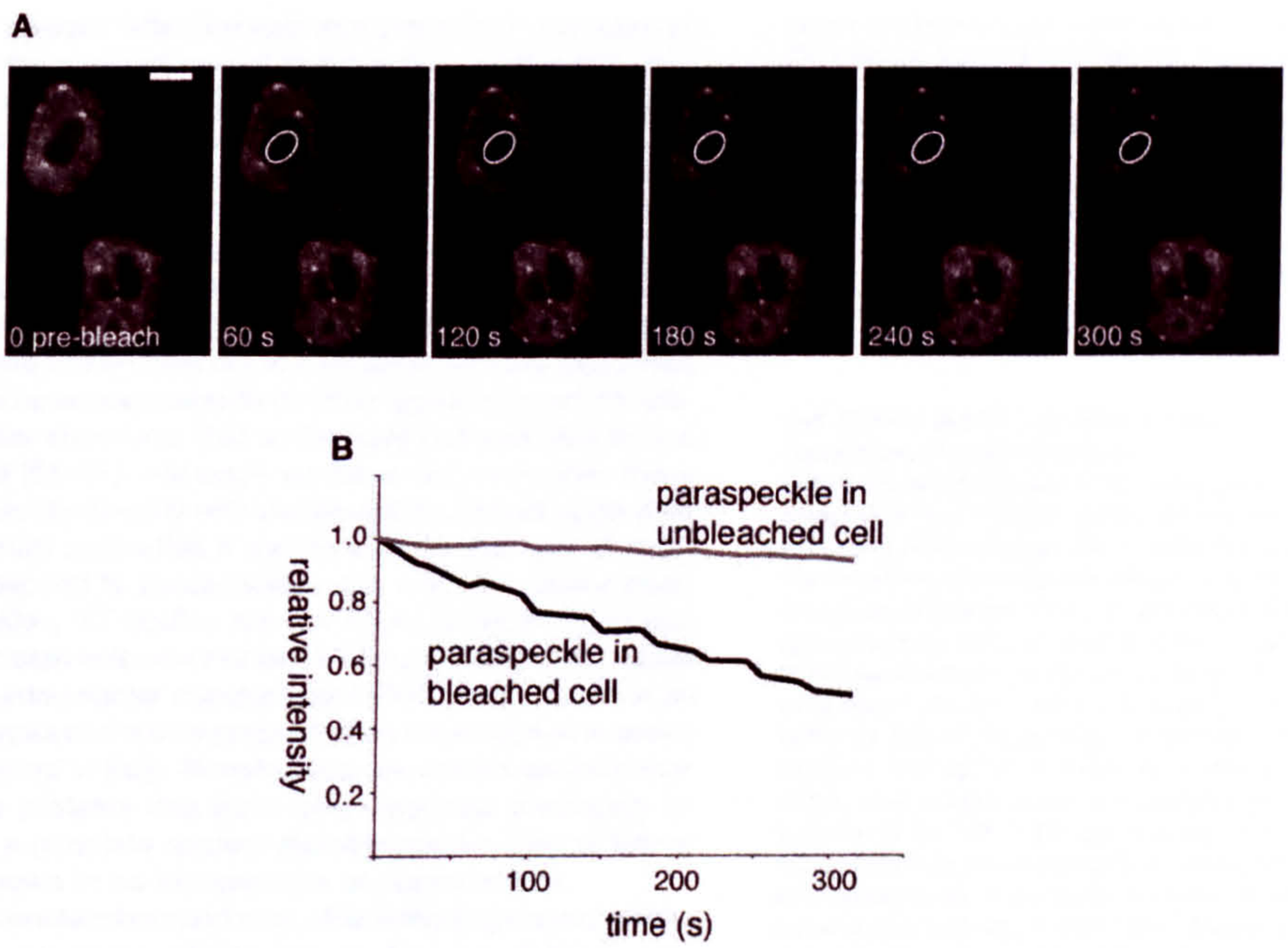


Figure 8. FLIP Experiments after Bleaching of Nucleoli in HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ Cells Demonstrate that YFP-PSP1 Cycles between the Nucleolus and Nucleoplasmic Structures

(A) Following the acquisition of a prebleach image, a HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cell was periodically bleached inside the outlined area within the nucleolus. Images were collected before and after bleaching. The images shown here are time points occurring every minute for a 5 min experiment. The repeated bleach pulses do not affect a neighboring cell nucleus. Scale bar, 5 μm.

(B) Quantitation of changes in fluorescence intensity within selected paraspeckles in the bleached and unbleached cells (shown in [A]) over the course of a 5 min experiment.

ribosome biogenesis that associate with both nucleoli and paraspeckles. This may reflect that PSP1, PSP2, and p54/nrb have hitherto unknown roles in ribosome assembly or function or that the nucleolus is involved in other nuclear functions. Data from the proteomic analysis of purified nucleoli also raised the possibility of additional nucleolar functions [32]. PSP1 is a novel protein whose cellular function is not known. In the case of p54/nrb, previous studies have implicated it in multiple processes, including transcriptional regulation, splicing, apoptosis, and dsRNA processing [45–49]. In particular, p54/nrb has recently been implicated, in partnership with other proteins, in cotranscriptional control through an interaction with steroid hormone nuclear receptors [50]. Interestingly, despite being uncharacterized at the time of its identification in the nucleolar proteome, PSP2 has recently been identified independently and reported to be a "coactivator activator" (CoAA) that coactivates transcription by association with the thyroid hormone receptor binding protein [51]. These related functions of p54/nrb and PSP2, in combination with the data showing that paraspeckle proteins accumulate at nucleoli when transcription is inhibited, suggest a possible role for the paraspeckle proteins in transcriptional control and also implicate the nucleolus in these processes.

The fact that all three paraspeckle proteins contain related RNA binding domains suggests that their function may be connected to RNA metabolism. It will be interesting, therefore, to determine which RNA substrates PSP1 and PSP2 can interact with and to see whether any other protein components of the paraspeckles contain similar RNA binding motifs. However, we note that the PSP1 RNA binding motif is unlikely to be sufficient to target proteins to paraspeckles, because our preliminary data indicate that the RNA binding region of PSP1 alone is unable to cause accumulation of YFP in paraspeckles (data not shown). Both PSP1 and p54/nrb also show homology to the RNA binding and carboxy-terminal regions of the splicing factor PSF [37]. Like the paraspeckle proteins, PSF was identified by MS as one of the factors that accumulated in nucleoli after Actinomycin D treatment [32]. This is consistent with the recent report that GFP-tagged PSF relocalizes predominantly to perinucleolar clusters following Actinomycin D treatment [52]. However, unlike the paraspeckle proteins, PSF was reported to colocalize with splicing factors in the nucleus when transcription is active [37, 52]. Proteins in the PSP1 family may have roles in various aspects of nuclear organization concerned with RNA processing. For example, p54/nrb was recently isolated in a complex with PSF and matrin 3 (a putative component of the nuclear matrix) that bound and retained double-stranded inosine-containing RNA inside the nucleus [49]. Furthermore, a *C. tentans* protein, hrp65, which has ~36% identity with both the human PSP1 and p54/nrb proteins, localizes to fine fibers

that interact with Balbiani ring pre-mRNP particles inside the nucleoplasm of insect cells, as judged by electron microscopy [53]. Interestingly, hrp65 structures do not associate with pre-mRNA cotranscriptionally. Considering that paraspeckles are distinct from sites of Br-UTP incorporation in mammalian cell nuclei, it is possible these hrp65 fibers represent insect cell structures related to paraspeckles.

In addition to the major classes of subnuclear bodies that are shown here not to colocalize with paraspeckles, there have been reports of other types of punctate subnuclear structures that so far have been studied in less detail [54–61]. Although we have not compared these structures directly with paraspeckles, based upon their reported properties it seems unlikely that any of them correspond to paraspeckles. For example, unlike paraspeckles, YT bodies are not found in MCF7 cells and colocalize with sites of BrUTP incorporation [61], while the perinucleolar compartment (PNC) is not found in all cell types and is only present when transcription is active (reviewed in [62]). Nonetheless, we cannot exclude that some proteins that have been reported previously to have a punctate nuclear staining pattern may in future be shown to be components of paraspeckles.

It is unclear how and why, after inhibiting transcription, some nuclear proteins segregate into distinct perinucleolar cap structures (as seen in Figure 7D). Interestingly, the paraspeckle proteins all reside within the same nucleolar cap (and appear excluded from the coilin or fibrillarin caps). Further to this, we have observed that the DEAD box factors p68 and p72 (also found enriched in nucleoli isolated from cells treated with Actinomycin D [32]) also colocalize with PSP1 in perinucleolar caps following treatment of cells with Actinomycin D (data not shown, see Supplementary Material). Thus, p68 and p72 may be either functionally or biochemically related to the paraspeckle proteins.

In conclusion, we have identified a novel nucleoplasmic structure, termed paraspeckles, detected in all human cell types analyzed and shown to contain at least three separate proteins. These proteins appear to cycle continually between paraspeckles and the nucleolus, but all three accumulate in perinucleolar caps when transcription is inhibited. The prevalence of paraspeckles and their dynamic relationship to the nucleolus provides a new challenge for the ongoing effort to understand the functional complexity of nuclear structure.

Experimental Procedures

Plasmid Constructs and Antibody Production

The two main species of ESTs for PSP1 and thus the most often expressed forms corresponded to the two different PSP1 isoforms (PSP1-β residues 1–393) and (PSP1-α residues 1–523). To clone these, PSP1-β was amplified by PCR using the EST IMAGE 3634993 (UK HGMP Resource Centre), and PSP1-α was amplified using the overlapping templates EST IMAGE 2440134 and IMAGE 3634993. In both cases, the fragments were purified, digested with EcoRI and BamHI, and ligated into pEYFP-C1 (Clontech).

To make a myc-tagged fusion of PSP1, PSP1-β was amplified using IMAGE 3634993 as a template; the product was digested with BamHI and HindIII and subcloned into pSG9M [63]. YFP-p54/nrb was made by amplification of p54/nrb from a HeLa mRNA library (Clontech). The amplified product was digested with HindIII and BamHI and ligated into pEYFP-C1. YFP-PSP2 was made by amplification of the full-length cDNA constructed using the overlapping ESTs IMAGE 2350028 and 3505412. The resultant product was then digested with EcoRI and BamHI and ligated into pEYFP-C1.

For probing the multitissue Northern blot, a cDNA probe containing sequence corresponding to 107–702 bp of PSP1 was radiolabeled and purified prior to hybridization according to the manufacturer's instructions (Clontech).

An affinity-purified peptide antibody against PSP1 was generated in rabbit, using the sequence APPAPAPPEDHPDEEM (Eurogentec). This sequence was chosen as it lies within a region of PSP1 with little or no similarity with other proteins and is thus unique to PSP1.

Cell Culture and Transfection Assays, Generation of Stable Cell Line

HeLa cells, obtained from ATCC, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 100 U/ml penicillin and streptomycin (Life Technologies). For detection of fluorescent fusion proteins with fluorescence microscopy, cells were grown on coverslips and transfected with the appropriate construct, using Effectene reagent (Qiagen) according to the manufacturer's instructions. At 12–16 hr posttransfection, the cells were fixed for 7 min in 3.7% (w/v) paraformaldehyde in CSK buffer (10 mM PIPES [pH 6.8], 10 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 2 mM EDTA) at room temperature, mounted on to glass slides, and imaged under microscopes as described below. For both forms of YFP-PSP1 and the myc-PSP1 expression vectors, these were transient transfections. Thus, there was variation in the expression levels of the fusion proteins, resulting in variation in the number and intensity of PSP1 foci observed, with overexpressing cells showing numerous (>100) bright nuclear and cytoplasmic foci and with low expressors resembling the anti-PSP1 antiserum staining pattern. Thus, we concluded that the low-expressing pattern reflected the genuine localization of PSP1.

The YFP-PSP1 stable cell lines were generated essentially as described in [64]. These stable cells showed a pattern of YFP-PSP1 localization that resembled the low-expressing transiently transfected cells.

Drug treatments were carried out as follows: Actinomycin D (1 μg/ml) (Sigma) or DRB (100 μM) (Calbiochem) in DMEM incubated for 4 hr followed by fixation and fluorescence microscopy.

Preparation of Cell Lysates and Immunoblotting

Confluent 10 cm dishes of cells were washed with PBS and lysed using 1× LDS sample buffer (Novex). Lysates were passed through a Qiashredder (Qiagen) before being denatured using 100 mM DTT and heating at 70°C, then electrophoresed on a 1 mm, 4%–12% tris-glycine polyacrylamide gel (Novex) with MOPS buffer (Novex). Proteins were subsequently transferred onto nitrocellulose membrane (Schleicher and Schuell) using a submarine system (Novex) and buffer containing 12 mM Tris, 100 mM glycine, and 20% methanol. Following blocking with 5% milk powder in PBS + 0.05% Tween 20, the membranes were incubated with rabbit anti-PSP1 antibody (1:2000 dilution) or mouse anti-GFP antibody (Roche, 1:1000 dilution) detected using rabbit HRP conjugate (1:2000 dilution) or mouse HRP conjugate (1:5000 dilution) (both Pierce Chemical Co.) in PBS containing 5% milk powder and 0.05% Tween 20, and detected via chemiluminescence with ECL plus (Amersham Pharmacia Biotech).

Salt and detergent extraction of HeLa and HeLa$^{YFP\text{-}PSP1\text{-}\beta}$ cells was performed as described in [65]. Approximately equal amounts of protein were loaded in each lane, as estimated by Ponceau staining of proteins transferred to nitrocellulose membrane in pilot experiments. Samples were then subjected to immunoblotting as described above, with anti-PSP1 antiserum and anti rabbit HRP conjugate.

Immunofluorescence Microscopy

Cells were washed in PBS and fixed (as above) at room temperature. Permeabilization was performed with 1% Triton X-100 in PBS for 15 min at room temperature. Immunofluorescence staining was carried out using standard techniques. Antibodies used were anti-PSP1 rabbit peptide antibody (dilution 1:50), anti-fibrillarin monoclonal 72b9 (dilution 1:10, [66]), anti-myc monoclonal 9E10 (dilution 1:100,

[67]), anti-SC35 monoclonal (dilution 1:1000, Sigma), anti-2,2,7-trimethy-guanosine monoclonal (dilution 1:10, Calbiochem), anti-p80 coilin monoclonal 5P10 (dilution 1:10), anti-PML monoclonal PGM3 (dilution 1:10, Santa Cruz), anti-SMN monoclonal (dilution 1:50, Transduction Laboratories), and FITC and TRITC conjugated secondary antibodies (Jackson Laboratories). Transcription sites were visualized in live cells using previously described methods [68], and the resultant Br-UTP-containing mRNA was detected using anti-Bromo-dUTP antibody (dilution 1:5, Roche). Prior to mounting on slides, coverslips were sometimes soaked in DAPI (1 μM in water for 1 min) to stain DNA and/or followed by soaking in Pyronin Y (0.66 mM in water for 2 s) to stain RNA. Cells were mounted in either Mowiol/Dabco (for the confocal microscope) or 0.5% p-phenylenediamine in 20 mM Tris (pH 8.8), 90% glycerol (for the deltavision system).

Fluorescence microscopy of fixed cells was carried out using a 40× NA 1.3, 63× NA 1.4, or a 100× NA 1.4 Plan-Apochromat objective. Three-dimensional images and sections were recorded either on a LSM410 confocal microscope (Zeiss) or on a Zeiss DeltaVision Restoration microscope (Applied Precision, Inc.). Images presented here are projections of the entire nuclear fluorescence, unless otherwise stated.

Observation of Live Cells and Photobleaching Analysis

For live cell imaging, cells were grown on 42 mm glass coverslips (no. 1; Helmut Sauer) in medium containing 200 μg/ml G418. Cells were maintained at 37°C by use of a closed perfusion chamber (Bachofer) in DMEM media without phenol red (Life Technologies) but with the addition of 20 mM HEPES. Images were collected using the 100× NA 1.4 Plan-Apochromat objective on the DeltaVision microscope. For each nucleus, 20 to 30 optical sections were recorded. The Hg lamp was attenuated with a 0.5 OD neutral density filter, and images were recorded using a binning of 3 × 3. Images were recorded every 10 min over a time period of 3 hr. Time-lapse images were viewed as 3D maximum intensity projections of each time point (SoftWoRx, Applied Precision, Inc.).

Photobleaching experiments were carried out on a ZEISS 510 confocal laser scanning microscope equipped with an argon-krypton laser (ZEISS). The 488 nm laser and a 63× plan Apolens with a 1.4 NA were used. A laser power of 2.5% was used in image acquisitions, and 25% was used for photobleaching. An area of 16 × 16 pixels was bleached with an iteration of 250 (duration of bleach was 3 s). An image was collected before and after every bleaching event, with 30 s intervals between each bleaching event.

Supplementary Material

The Supplementary Material contains two figures: (1) a diagram detailing the gene structure of human PSP1, including a representation of which exons are included in the major splice forms of PSP1, and (2) fluorescence micrographs showing that in HeLa cells treated with the transcription inhibitor Actinomycin D, YFP-p68 localizes to the same perinucleolar cap structure as PSP1. The Supplementary Material can be found at http://images.cellpress.com/supmat/supmatin.htm.

Acknowledgments

A.I.L. is a Wellcome Trust Principal Research Fellow and is funded by a Wellcome Trust Programme grant. A.H.F. is funded by a Wellcome Trust International Travelling Fellowship; A.K.L.L. is funded by a Croucher studentship; Y.W.L. is funded by a Croucher postdoctoral fellowship; and C.E.L. is funded by the Wellcome Trust. We would like to thank Dr. K. Collins for the 1787htert cell line. Work in M.M.'s laboratory is funded by a Danish National Research Foundation grant to the Centre for Experimental BioInformatics. We thank Dr. B. Frenguelli and Dr. C. Connolly from Ninewells Hospital, Dundee, for the use of the LSM510.

Received: October 15, 2001
Revised: November 14, 2001
Accepted: November 14, 2001
Published: January 8, 2002

References

1. Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. 2, 292–301.
2. Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science 280, 547–553.
3. Pederson, T. (2001). Protein mobility within the nucleus—what are the right moves? Cell 104, 635–638.
4. Misteli, T. (2001). Protein dynamics: implications for nuclear architecture and gene expression. Science 291, 843–847.
5. Matera, A.G. (1999). Nuclear bodies: multifaceted subdomains of the interchromatin space. Trends Cell Biol. 9, 302–309.
6. Weis, K., Rambaud, S., Lavau, C., Jansen, J., Carvalho, T., Carmo-Fonseca, M., Lamond, A., and Dejean, A. (1994). Retinoic acid regulates aberrant nuclear localization of PML-RAR alpha in acute promyelocytic leukemia cells. Cell 76, 345–356.
7. Dyck, J.A., Maul, G.G., Miller, W.H., Jr., Chen, J.D., Kakizuka, A., and Evans, R.M. (1994). A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. Cell 76, 333–343.
8. Koken, M.H., Puvion-Dutilleul, F., Guillemin, M.C., Viron, A., Linares-Cruz, G., Stuurman, N., de Jong, L., Szostecki, C., Calvo, F., Chomienne, C., et al. (1994). The t(15;17) translocation alters a nuclear body in a retinoic acid-reversible fashion. EMBO J. 13, 1073–1083.
9. Liu, Q., and Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. EMBO J. 15, 3555–3565.
10. Bridge, E., Xia, D.X., Carmo-Fonseca, M., Cardinali, B., Lamond, A.I., and Pettersson, U. (1995). Dynamic organization of splicing factors in adenovirus-infected cells. J. Virol. 69, 281–290.
11. Phelan, A., Carmo-Fonseca, M., McLaughlan, J., Lamond, A.I., and Clements, J.B. (1993). A herpes simplex virus type 1 immediate-early gene product, IE63, regulates small nuclear ribonucleoprotein distribution. Proc. Natl. Acad. Sci. USA 90, 9056–9060.
12. Santama, N., Dotti, C.G., and Lamond, A.I. (1996). Neuronal differentiation in the rat hippocampus involves a stage-specific reorganization of subnuclear structure both in vivo and in vitro. Eur. J. Neurosci. 8, 892–905.
13. Dahm, R., Gribbon, C., Quinlan, R.A., and Prescott, A.R. (1998). Changes in the nucleolar and coiled body compartments precede lamina and chromatin reorganization during fibre cell denucleation in the bovine lens. Eur. J. Cell Biol. 75, 237–246.
14. Antoniou, M., Carmo-Fonseca, M., Ferreira, J., and Lamond, A.I. (1993). Nuclear organization of splicing snRNPs during differentiation of murine erythroleukemia cells in vitro. J. Cell Biol. 123, 1055–1068.
15. Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. Nature 404, 604–609.
16. Snaar, S., Wiesmeijer, K., Jochemsen, A.G., Tanke, H.J., and Dirks, R.W. (2000). Mutational analysis of fibrillarin and its mobility in living human cells. J. Cell Biol. 151, 653–662.
17. Chen, D., and Huang, S. (2001). Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. J. Cell Biol. 153, 169–176.
18. Stenoien, D.L., Patel, K., Mancini, M.G., Dutertre, M., Smith, C.L., O'Malley, B.W., and Mancini, M.A. (2001). FRAP reveals that mobility of oestrogen receptor-alpha is ligand- and proteasome-dependent. Nat. Cell Biol. 3, 15–23.
19. Spector, D.L., Fu, X.D., and Maniatis, T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. EMBO J. 10, 3467–3481.
20. Carmo-Fonseca, M., Pepperkok, R., Carvalho, M.T., and Lamond, A.I. (1992). Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. J. Cell Biol. 117, 1–14.
21. Carter, K.C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J.A., Fay, F.S., and Lawrence, J.B. (1993). A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. Science 259, 1330–1335.
22. Misteli, T., and Spector, D.L. (1998). The cellular organization of gene expression. Curr. Opin. Cell Biol. 10, 323–331.

23. Dirks, R.W., Hattinger, C.M., Molenaar, C., and Snaar, S.P. (1999). Synthesis, processing, and transport of RNA within the three-dimensional context of the cell nucleus. Crit. Rev. Eukaryot. Gene Expr. *9*, 191–201.
24. Sleeman, J.E., and Lamond, A.I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. Curr. Biol. *9*, 1065–1074.
25. Speckmann, W., Narayanan, A., Terns, R., and Terns, M.P. (1999). Nuclear retention elements of U3 small nucleolar RNA. Mol. Cell. Biol. *19*, 8412–8421.
26. Yu, Y.T., Shu, M.D., Narayanan, A., Terns, R.M., Terns, M.P., and Steitz, J.A. (2001). Internal modification of u2 small nuclear (sn)rna occurs in nucleoli of xenopus oocytes. J. Cell Biol. *152*, 1279–1288.
27. Gall, J.G., Bellini, M., Wu, Z., and Murphy, C. (1999). Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. Mol. Biol. Cell *10*, 4385–4402.
28. Frey, M.R., Bailey, A.D., Weiner, A.M., and Matera, A.G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. Curr. Biol. *9*, 126–135.
29. Pederson, T. (1998). The plurifunctional nucleolus. Nucleic Acids Res. *26*, 3871–3876.
30. Olson, M.O., Dundr, M., and Szebeni, A. (2000). The nucleolus: an old factory with unexpected capabilities. Trends Cell Biol. *10*, 189–196.
31. Carmo-Fonseca, M., Mendes-Soares, L., and Campos, I. (2000). To be or not to be in the nucleolus. Nat. Cell Biol. *2*, E107–112.
32. Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K.L., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. Curr. Biol. *12*, 1–11.
33. Dong, B., Horowitz, D.S., Kobayashi, R., and Krainer, A.R. (1993). Purification and cDNA cloning of HeLa cell p54nrb, a nuclear protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and Drosophila NONA/BJ6. Nucleic Acids Res. *21*, 4085–4092.
34. Yang, Y.S., Hanke, J.H., Carayannopoulos, L., Craft, C.M., Capra, J.D., and Tucker, P.W. (1993). NonO, a non-POU-domain-containing, octamer-binding protein, is the mammalian homolog of Drosophila nonAdiss. Mol. Cell. Biol. *13*, 5593–5603.
35. Visa, N., Puvion-Dutilleul, F., Bachellerie, J.P., and Puvion, E. (1993). Intranuclear distribution of U1 and U2 snRNAs visualized by high resolution in situ hybridization: revelation of a novel compartment containing U1 but not U2 snRNA in HeLa cells. Eur. J. Cell Biol. *60*, 308–321.
36. Stuurman, N., de Graaf, A., Floore, A., Josso, A., Humbel, B., de Jong, L., and van Driel, R. (1992). A monoclonal antibody recognizing nuclear matrix-associated nuclear bodies. J. Cell Sci. *101*, 773–784.
37. Patton, J.G., Porro, E.B., Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. Genes Dev. *7*, 393–406.
38. Lamm, G.M., Nicol, S.M., Fuller-Pace, F.V., and Lamond, A.I. (1996). p72: a human nuclear DEAD box protein highly related to p68. Nucleic Acids Res. *24*, 3739–3747.
39. Andrade, L.E., Chan, E.K., Raska, I., Peebles, C.L., Roos, G., and Tan, E.M. (1991). Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. J. Exp. Med. *173*, 1407–1419.
40. Matunis, M.J., Michael, W.M., and Dreyfuss, G. (1992). Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. Mol. Cell. Biol. *12*, 164–171.
41. Raska, I., Ochs, R.L., Andrade, L.E., Chan, E.K., Burlingame, R., Peebles, C., Gruol, D., and Tan, E.M. (1990). Association between the nucleolus and the coiled body. J. Struct. Biol. *104*, 120–127.
42. Ochs, R.L., Lischwe, M.A., Spohn, W.H., and Busch, H. (1985). Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. Biol. Cell *54*, 123–133.
43. Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S., and Spector, D.L. (1999). Purification and biochemical characterization of interchromatin granule clusters. EMBO J. *18*, 4308–4320.
44. Visintin, R., and Amon, A. (2000). The nucleolus: the magician's hat for cell cycle tricks. Curr. Opin. Cell Biol. *12*, 752.
45. Thiede, B., Dimmler, C., Siejak, F., and Rudel, T. (2001). Predominant identification of RNA-binding proteins in Fas-induced apoptosis by proteome analysis. J. Biol. Chem. *276*, 26044–26050.
46. Hallier, M., Tavitian, A., and Moreau-Gachelin, F. (1996). The transcription factor Spi-1/PU.1 binds RNA and interferes with the RNA-binding protein p54nrb. J. Biol. Chem. *271*, 11177–11181.
47. Basu, A., Dong, B., Krainer, A.R., and Howe, C.C. (1997). The intracisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54nrb/NonO. Mol. Cell. Biol. *17*, 677–686.
48. Yang, Y.S., Yang, M.C., Tucker, P.W., and Capra, J.D. (1997). NonO enhances the association of many DNA-binding proteins to their targets. Nucleic Acids Res. *25*, 2284–2292.
49. Zhang, Z., and Carmichael, G.G. (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell *106*, 465–475.
50. Mathur, M., Tucker, P.W., and Samuels, H.H. (2001). PSF is a novel corepressor that mediates its effect through Sin3A and the DNA binding domain of nuclear hormone receptors. Mol. Cell. Biol. *21*, 2298–2311.
51. Iwasaki, T., Chin, W.W., and Ko, L. (2001). Identification and characterization of rrm-containing coactivator activator (coaa) as trbp-interacting protein, and its splice variant as a coactivator modulator (coam). J. Biol. Chem. *276*, 33375–33383.
52. Dye, B.T., and Patton, J.G. (2001). An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles. Exp. Cell Res. *263*, 131–144.
53. Miralles, F., Ofverstedt, L.G., Sabri, N., Aissouni, Y., Hellman, U., Skoglund, U., and Visa, N. (2000). Electron tomography reveals posttranscriptional binding of pre-mRNPs to specific fibers in the nucleoplasm. J. Cell Biol. *148*, 271–282.
54. Chiodi, I., Biggiogera, M., Denegri, M., Corioni, M., Weighardt, F., Cobianchi, F., Riva, S., and Biamonti, G. (2000). Structure and dynamics of hnRNP-labelled nuclear bodies induced by stress treatments. J. Cell Sci. *113*, 4043–4053.
55. Weighardt, F., Cobianchi, F., Cartegni, L., Chiodi, I., Villa, A., Riva, S., and Biamonti, G. (1999). A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock. J. Cell Sci. *112*, 1465–1476.
56. Prasanth, K.V., Rajendra, T.K., Lal, A.K., and Lakhotia, S.C. (2000). Omega speckles—a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in Drosophila. J. Cell Sci. *113*, 3485–3497.
57. Elefanty, A.G., Antoniou, M., Custodio, N., Carmo-Fonseca, M., and Grosveld, F.G. (1996). GATA transcription factors associate with a novel class of nuclear bodies in erythroblasts and megakaryocytes. EMBO J. *15*, 319–333.
58. Saurin, A.J., Shiels, C., Williamson, J., Satijn, D.P., Otte, A.P., Sheer, D., and Freemont, P.S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J. Cell Biol. *142*, 887–898.
59. Cotto, J., Fox, S., and Morimoto, R. (1997). HSF1 granules: a novel stress-induced nuclear compartment of human cells. J. Cell Sci. *110*, 2925–2934.
60. Chen, T., Boisvert, F.M., Bazett-Jones, D.P., and Richard, S. (1999). A role for the GSG domain in localizing Sam68 to novel nuclear structures in cancer cell lines. Mol. Biol. Cell *10*, 3015–3033.
61. Nayler, O., Hartmann, A.M., and Stamm, S. (2000). The ER repeat protein YT521-B localizes to a novel subnuclear compartment. J. Cell Biol. *150*, 949–962.
62. Huang, S. (2000). Review: perinucleolar structures. J. Struct. Biol. *129*, 233–240.
63. Green, S., Issemann, I., and Sheer, E. (1988). A versatile in vivo and in vitro eukaryotic expression vector for protein engineering. Nucleic Acids Res. *16*, 369.
64. Sleeman, J., Ajuh, P., and Lamond, A.I. (2001). snRNP protein

expression enhances the formation of Cajal bodies containing p80-Coilin and SMN. J. Cell Sci., in press.
65. Platani, M., Goldberg, I., Swedlow, J.R., and Lamond, A.I. (2000). In vivo analysis of Cajal body movement, separation, and joining in live human cells. J. Cell Biol. *151*, 1561–1574.
66. Reimer, G., Raska, I., Tan, E.M., and Scheer, U. (1987). Human autoantibodies: probes for nucleolus structure and function. Virchows Arch. B Cell Pathol. Incl. Mol. Pathol. *54*, 131–143.
67. Evan, G.I., Lewis, G.K., Ramsay, G., and Bishop, J.M. (1985). Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. Mol. Cell. Biol. *5*, 3610–3616.
68. Masson, C., Bouniol, C., Fomproix, N., Szollosi, S., Debey, P., and Hernandez-Verdun, D. (1996). Conditions favouring RNA polymerase I transcription in permeabilized cells. Exp. Cell Res. *226*, 114–125.

Accession Numbers

The GenBank accession numbers for PSP1-α and PSP1-β have been deposited as AF448795 and AF449627, respectively.

Appendix II:

Supplementary Information

CD Content

CD Contents

The Nucleolar Protein Database and relevant programming scripts

Table III-S1	Nucleolar Proteome Motif Distribution
Table III-S2	Nucleolar Profiling Details and *in silico* classification calculation procedure
Full Table III-3	Nucleolar Protein Database (Microsoft Excel format)
Movie IV-5	Timing of DFC and GC breakdown and assembly
Movie IV-6	Timing of FC and DFC breakdown
Movie IV-9	Nucleolar Reassembly
Movie IV-10	Same genetic materials, different structures
Movie IV-11	Chromatin and the Nucleolus

The supplementary information CD is accompanied at the back of this page.